

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

2

JC675 U.S. PTO
09/435034
11/05/99



Applicant(s): Itaru NISHIZAWA, ET AL.

Serial No.:

Filed: November 5, 1999

Title: DATA WAREHOUSE SYSTEM AND QUERY PROCESSING
METHOD USED IN THE SYSTEM AND DATA COLLECTING
METHOD AND APPARATUS FOR THE METHOD, AND
CHARGING METHOD AND APPARATUS IN THE SYSTEM

Group:

LETTER CLAIMING RIGHT OF PRIORITY

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

November 5, 1999

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the
applicant(s) hereby claim(s) the right of priority based on
Japanese Patent Application No.(s) 10-320203 filed November 11,
1998.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP


Carl I. Brundidge
Registration No. 29,621

CIB/rdh
Attachment
(703) 312-6600

日本国特許庁

PATENT OFFICE
JAPANESE GOVERNMENT

— MK 319800888

JC675 U.S. PRO
09/435034
11/05/99

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application:

1998年11月11日

出願番号
Application Number:

平成10年特許願第320203号

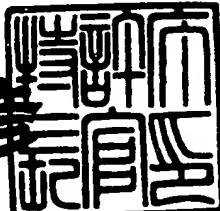
出願人
Applicant(s):

株式会社日立製作所

1999年10月15日

特許庁長官
Commissioner,
Patent Office

近藤 隆



【書類名】 特許願

【整理番号】 H98008881A

【提出日】 平成10年11月11日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 12/00

【発明の名称】 データウェアハウスシステムとそこで用いられる問合せ処理方法及びそのためのデータ収集方法と装置及び課金システム

【請求項の数】 34

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

【氏名】 西澤 格

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

【氏名】 藤原 真二

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

【氏名】 牛嶋 一智

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

【氏名】 猪原 茂和

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100068504

【弁理士】

【氏名又は名称】 小川 勝男

【電話番号】 03-3212-1111

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 データウェアハウスシステムとそこで用いられる問合せ処理方法及びそのためのデータ収集方法と装置及び課金システム

【特許請求の範囲】

【請求項1】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、

該データ収集手段で収集されたデータを格納するデータベースと、

該データ収集手段を介して各クライアントを所望のサーバに接続するネットワークとを備え、

該データ収集手段は、

少なくとも1個のクライアントを管理するクライアント管理部と、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、

該問合せ解析部での解析結果に基づいて少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、

解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と該レプリカを格納する前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、前記サーバは、

少なくとも1個の前記データ収集手段から送信されてきた問合せ解析結果を受信する通信制御部と、

前記サーバ側のデータベースを検索して、受信した問合せ解析結果に対する応答を作成する問合せ処理部とを有する、

データウェアハウスシステム。

【請求項2】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機シス

テム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、

該データ収集手段で収集されたデータを格納するデータベースとを備え、

該データ収集手段は、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、
解析した作成要求に従ってデータの共有が可能なクライアントをグループ化し管
理するクライアント管理部と、

解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記デ
ータベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用
いて該レプリカを管理するレプリカ作成管理部とを有する、

データウェアハウスシステム。

【請求項3】

前記レプリカ作成管理部は、前記クライアントからのレプリカ作成要求を受け
付けるためのプログラムを、ネットワークを介して受け取ったソフトウェア部品
群を組み合わせて動的に作成するプログラム作成部を含む請求項1記載のデータ
ウェアハウスシステム。

【請求項4】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機シス
テム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、

該データ収集手段で収集されたデータを格納するデータベースとを備え、

該データ収集手段は、

少なくとも1個のクライアントを管理するクライアント管理部と、

該クライアントからの要求を解析する問合せ解析部と、

該問合せ解析部での解析結果に基づいて少なくとも1個の前記サーバへのアクセ

手順を決定する通信制御部と、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、
解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、

前記サーバは、

少なくとも1個の前記データ収集手段から送信されてきた問合せ解析結果を受信する通信制御部と、

前記サーバ側のデータベースを検索し、受信した問合せ解析結果に対する応答を作成する問合せ処理部と、

前記データ収集手段に配達するデータの領域条件と、品質条件と、配達先と、配達方法とを含む配達データ管理テーブルと、

該テーブルに基づいて前記配達データを管理する配達データ管理部とを有する、データウェアハウスシステム。

【請求項5】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、

該データ収集手段で収集されたデータを格納するデータベースとを備え、

該データ収集手段は、

少なくとも1個のクライアントを管理するクライアント管理部と、

該クライアントからの要求を解析する問合せ解析部と、

該問合せ解析部での解析結果に基づいて少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、
解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用

いて該レプリカを管理するレプリカ作成管理部とを有し、
前記サーバは、
少なくとも1個の前記データ収集手段から送信されてきた問合せ解析結果を受信
する通信制御部と、
前記サーバ側のデータベースを検索し、受信した問合せ解析結果に対する応答を
作成する問合せ処理部と、
前記データ収集手段に配送する配送データの領域条件と品質条件と、該配送データ
の配送先と配送方法とを含む配送データ管理テーブルと、
該管理テーブルに基づいて前記配送データを管理する配送データ管理部と、
前記サーバの負荷情報を管理し、前記配送データに対して新たに受け付けた配送
要求を処理するときのサーバ負荷を見積る負荷管理部とを有する、
データウェアハウスシステム。

【請求項6】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機シス
テム（以下、クライアント）と、
それぞれにデータベースを備え該クライアントからのアクセス要求に従って該デ
ータベースを検索するサーバ計算機システム（以下、サーバ）と、
ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、
該データ収集手段で収集されたデータを格納するデータベースとを備え、
該データ収集手段は、
少なくとも1個のクライアントを管理するクライアント管理部と、
該クライアントからの要求を解析する問合せ解析部と、
該問合せ解析部で解析された問合せを処理する問合せ処理部と、
該問合せ解析部での解析結果に基づいて少なくとも1個の前記サーバへのアクセ
ス手順を決定する通信制御部と、
該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、
解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記デ
ータベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用
いて該レプリカを管理するレプリカ作成管理部とを有し、

前記問合せ処理部は、クライアントがアクセスするデータの一部あるいは全部を、クライアントから与えられるレプリカ作成要求に基づいて、クライアントの問合せ処理に先んじてレプリカを事前生成するレプリカ事前生成部を含む、データウェアハウスシステム。

【請求項7】

前記データ収集手段は、前記レプリカ作成管理部でレプリカを作成する際に、前記レプリカ作成要求と、該データ収集手段で利用可能な計算機資源および前記サーバで利用可能な計算機資源に基づき、該サーバと交渉処理を行うことによってレプリカ作成対象のデータを決定するデータ収集手段交渉処理部を含む請求項5記載のデータウェアハウスシステム。

【請求項8】

前記サーバは、前記負荷管理部によって見積もられるサーバ負荷情報と、該サーバで利用可能な計算機資源および前記データ収集手段で利用可能な計算機資源に基づき、前記データ収集手段と交渉処理を行うサーバ交渉処理部を有する請求項6記載のデータウェアハウスシステム。

【請求項9】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、
それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、
ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、
該データ収集手段で収集されたデータを格納するデータベースとを備え、
該データ収集手段は、
少なくとも1個のクライアントを管理するクライアント管理部と、
該クライアントからの要求を解析する問合せ解析部と、
該問合せ解析部での解析結果に基づいて少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、
該クライアントからのレプリカ作成要求が、該クライアントがアクセスするデータの範囲、データの精度、データの鮮度、データの優先度をそれぞれ指定する条

件のうちの少なくとも1個の条件を含むレプリカ作成要求を解析するレプリカ作成要求解析部と、

解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有する、
データウェアハウスシステム。

【請求項10】

前記サーバは、データの精度を指定する条件を含むレプリカ作成要求に対し、取り扱うデータの型に応じて、以下のいずれかの処理によってデータを抽出するデータ抽出部を有する請求項9記載のデータウェアハウスシステム。

- (1) 関係データベースのレコードデータに対しては、サンプリング、もしくはフィールド切り出しによる、
- (2) 文書データに対しては、要約作成、もしくはキーワード切り出しによる、
- (3) 静止画像データに対しては、可逆圧縮、非可逆圧縮、輪郭抽出、色数削減、解像度削減、サイズ縮小のいずれかによる、
- (4) 動画像データに対しては、フレーム数削減、もしくはフレーム内画像圧縮による、
- (5) 音声データに対しては、サンプリングレート削減、もしくは音質調整、もしくは文字データへの変換による。

【請求項11】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、該データ収集手段で収集されたデータを格納するデータベースとを備え、該データ収集手段は、

少なくとも1個のクライアントを管理するクライアント管理部と、該クライアントからの要求を解析する問合せ解析部と、

該問合せ解析部での解析結果に基づいて少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、

該レプリカ作成管理部は、前記レプリカ作成要求と、前記データ収集手段で利用可能な計算機資源および前記サーバで利用可能な計算機資源に基づき、該サーバと交渉を行うことによって、レプリカ作成対象のデータを決定するデータ収集手段交渉処理部を含み、

前記サーバは、

少なくとも1個の前記データ収集手段から送信されてきた問合せ解析結果を受信する通信制御部と、

該サーバで利用可能な計算機資源に基づき、前記データ収集手段との間でレプリカ作成条件に関する交渉を行うサーバ交渉処理部と、

前記サーバ側のデータベースを検索し、受信した問合せ解析結果に対する応答を作成する問合せ処理部とを有する、

データハウスシステム。

【請求項12】

前記サーバは、配送データのレプリカ作成要求を新たに受け付けて処理するときの該サーバの負荷を見積る負荷管理部を有する請求項11記載のデータハウスシステム。

【請求項13】

前記サーバ交渉処理部は、前記サーバで利用可能な計算機資源として前記負荷管理部によって見積られる該サーバの負荷情報を含むような計算機資源に基づき、前記データ収集手段との交渉を行う処理部を含む請求項12記載のデータハウスシステム。

【請求項14】

前記サーバ交渉処理部は、前記レプリカ作成要求に対して、受け付けるか、拒

否するか、条件付きで受け付けるかのいずれかの応答を返す応答処理部を含む請求項11記載のデータウェアハウスシステム。

【請求項15】

前記応答処理部は、条件付きで受け付ける応答に、前記レプリカ作成要求の条件を変更して作成した新条件を付与して該データ収集手段に転送する転送処理部を含む請求項14記載のデータウェアハウスシステム。

【請求項16】

前記データ収集手段交渉処理部は、前記応答に付与されて転送されてきた新条件を、クライアントに提示する処理部を含む請求項15記載のデータウェアハウスシステム。

【請求項17】

前記レプリカ作成管理部は、前記交渉処理の結果作成したレプリカのデータ領域条件とデータ品質条件、該レプリカの格納位置情報、および該レプリカの配送方法を、前記レプリカ管理テーブルに加える追加処理部を含む請求項11記載のデータウェアハウスシステム。

【請求項18】

前記配送データ管理部は、前記交渉処理の結果提供することを決定したデータの領域条件と品質条件、該データの提供先、および該データの配送方法を、前記配送データ管理テーブルに加える追加処理部を含む請求項11記載のデータウェアハウスシステム。

【請求項19】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、

該データ収集手段で収集されたデータを格納するデータベースと、

該データ収集手段を介して各クライアントを所望のサーバに接続するネットワークとを備え、

該データ収集手段は、

少なくとも1個のクライアントを管理するクライアント管理部と、

該クライアントからの要求を解析する問合せ解析部と、

該問合せ解析部で解析された問合せを処理する問合せ処理部と、

該問合せ解析部での解析結果に基づいて前記ネットワークで接続される少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、
解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、

前記サーバは、

少なくとも1個の前記データ収集手段から前記ネットワークを通して送信されてきた問合せ解析結果を受信する通信制御部と、

前記サーバ側のデータベースを検索し、受信した問合せ解析結果に対する応答を作成する問合せ処理部と、

を有し、さらに、

前記サーバとデータ収集手段間で、サーバ主導でデータを配達するか、データ収集手段主導でデータを配達するかの少なくとも一方を選択的に利用できるように制御する手段を備えた、

データウェアハウスシステム。

【請求項20】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、

該データ収集手段で収集されたデータを格納するデータベースと、

該データ収集手段を介して各クライアントを所望のサーバに接続するネットワークとを備え、

該データ収集手段は、

少なくとも1個のクライアントを管理するクライアント管理部と、

該クライアントからの要求を解析する問合せ解析部と、

該問合せ解析部での解析結果に基づいて前記ネットワークで接続される少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、

解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、

前記問合せ解析部は、

前記レプリカ管理テーブルを用いて前記問合せを処理すべきデータ収集手段を決定する決定処理部と、

該問合せ処理すべきデータ収集手段のレプリカを用いた問合せ処理が可能か否か判定する判定処理部と、

可能と判定された場合に、前記サーバへ問合せを転送することなく該レプリカを用いた問合せ処理を前記問合せ処理部に指示する処理と、

否と判定された場合に、前記サーバへ問合せを転送する転送処理部とを含む、データウェアハウスシステム。

【請求項21】

前記位置決定処理部は、前記問合せ解析部により問合せと前記レプリカ内容との問合せの同値関係もしくは包含関係を決定する処理部分と、該同値関係もしくは包含関係と前記レプリカ管理テーブルとを利用して前記位置を決定する処理部分とを含む請求項20記載のデータウェアハウスシステム。

【請求項22】

前記位置決定処理部は、前記問合せ解析部により問合せと前記レプリカ内容との同値関係もしくは包含関係を決定する処理部分と、該問合せと該レプリカ内容とが同値関係もしくは包含関係にあるか否かを判定する処理部分と、あると判定された場合には該問合せを処理する位置を該レプリカが存在する前記データ収集手段と決定する処理部分と、否と判定された場合には該問合せ処理する位置を前

記サーバと決定する処理部分とを含む請求項20記載のデータウェアハウスシステム。

【請求項23】

前記問合せと前記レプリカ内容の同値関係もしくは包含関係を決定する処理部分は、

Query Equivalent 判定方式を用いて同値関係を決定する部分と、Query Containment 判定方式を用いて包含関係を決定する部分とを含む請求項21記載のデータウェアハウスシステム。

【請求項24】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、

それぞれにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、

ユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、

該データ収集手段で収集されたデータを格納するデータベースと、

該データ収集手段を介して各クライアントを所望のサーバに接続するネットワークとを備え、

該データ収集手段は、

少なくとも1個のクライアントを管理するクライアント管理部と、

該クライアントからの要求を解析する問合せ解析部と、

該問合せ解析部での解析結果に基づいて前記ネットワークで接続される少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、

該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用い、前記問合せ処理時には前記協調するデータ収集手段間でレプリカを共用できるように該レプリカを管理するレプリカ作成管理部とを有し、

前記サーバは、

協調する少なくとも1個の前記データ収集手段から前記ネットワークを通して送

信されてきた問合せ解析結果を受信する通信制御部と、
前記サーバ側のデータベースを検索して、受信した問合せ解析結果に対する応答
を作成する問合せ処理部とを有する
データウェアハウスシステム。

【請求項25】

前記レプリカ作成要求解析部は、前記レプリカ作成要求がデータの鮮度を指定
する条件を有し、前記サーバでデータが更新されたときの更新時刻と現時刻との
差が所定値よりも小さいならば、現時刻でのデータを前記データ収集手段内のデ
ータとして扱う処理部を含む請求項9記載のデータウェアハウスシステム。

【請求項26】

データウェアハウスに用いられるデータ収集装置であって、
少なくとも1個のクライアントを管理し該クライアントの要求するデータを収集
するための
クライアント管理部と、
該クライアントからの問合せ要求を解析する問合せ解析部と、
該解析された問合せ要求に含まれるレプリカ作成要求を解析するレプリカ作成要
求解析部と、
解析された作成要求に従って作成されるレプリカを格納するレプリカ用データベ
ースと、
解析された作成要求に従ってレプリカを作成し、作成したレプリカの内容と該レ
プリカ用データベース上の格納位置とに関する情報を対応づけたレプリカ管理テ
ーブルを用いて該レプリカを管理するレプリカ作成管理部とを有する、
データ収集装置。

【請求項27】

前記レプリカ用データベースに、汎用のデータベースマネージメントシステム
を用いる請求項26記載のデータ収集装置。

【請求項28】

データウェアハウスに用いられるデータ収集方法であって、
少なくとも1個のクライアントを管理し該クライアントの要求するデータを収集

するための

クライアント管理処理と、

該クライアントからの問合せ要求を解析する問合せ解析処理と、

該問合せ要求に含まれるレプリカ作成要求を解析するレプリカ作成要求解析処理と、

解析された作成要求に従って作成されるレプリカを格納するレプリカ用データベースと、

解析された作成要求に従ってレプリカを作成し、作成したレプリカの内容と該レプリカ用データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有する、データ収集方法。

【請求項29】

データウェアハウスに用いられる問合せ処理方法であって、少なくとも1個のクライアントを管理し該クライアントの要求するデータを収集するための

クライアント管理処理と、

該クライアントからの問合せ要求を解析する問合せ解析処理と、

該問合せ要求に含まれるレプリカ作成要求を解析するレプリカ作成要求解析処理と、

解析された作成要求に従ってレプリカを作成し、作成したレプリカの内容と該レプリカを格納するデータベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理処理と、

前記解析された問合せ要求を、該レプリカ管理テーブルにより管理されたレプリカを用いて処理する問合せ処理とを有する、

問合せ処理方法。

【請求項30】

請求項28または29の方法を実行するためのコンピュータプログラムを記録した記録媒体。

【請求項31】

ユーザからの処理要求を受け付ける少なくとも1個のクライアント計算機システム（以下、クライアント）と、
それにデータベースを備え該クライアントからのアクセス要求に従って該データベースを検索するサーバ計算機システム（以下、サーバ）と、
各クライアントを所望のサーバに接続するネットワークと、
該ネットワークに接続されユーザの要求するデータを収集する少なくとも1個のデータ収集手段と、
該データ収集手段で収集されたデータを格納するデータベースとを備え、
該データ収集手段は、
少なくとも1個のクライアントを管理するクライアント管理部と、
該クライアントからの要求を解析する問合せ解析部と、
該問合せ解析部での解析結果に基づいて前記ネットワークで接続される少なくとも1個の前記サーバへのアクセス手順を決定する通信制御部と、
該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、
解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と前記データベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、
前記サーバは、
少なくとも1個の前記データ収集手段から前記ネットワークを通して送信されてきた問合せ解析結果を受信する通信制御部と、
前記サーバ側のデータベースを検索し、受信した問合せ解析結果に対する応答を作成する問合せ処理部とを有する、
データウェアハウスシステム。

【請求項32】

前記レプリカ作成管理部は、前記クライアントからのレプリカ作成要求を受け付けるためのプログラムを、ネットワークを介して受け取ったソフトウェア部品群を組み合わせて動的に作成するプログラム作成部を含む請求項31記載のデータウェアハウスシステム。

【請求項33】

データウエアハウスに用いられる課金システムであって、少なくとも1個のクライアントを管理し該クライアントの要求するデータを収集するクライアント管理部と、該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と該レプリカを格納するデータベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、該レプリカ作成管理部は、前記レプリカ作成要求と、前記データ収集手段で利用可能な計算機資源およびサーバ側で利用可能な計算機資源に基づき、該サーバと交渉を行うことによって、レプリカ作成対象のデータと対価としての料金とを決定するデータ収集側交渉処理部と、該サーバで利用可能な計算機資源に基づき、前記データ収集側との間で前記料金を含むレプリカ作成条件に関する交渉を行うサーバ側交渉処理部とを有する、課金システム。

【請求項34】

データウエアハウスに用いられる課金方法であって、少なくとも1個のクライアントを管理し該クライアントの要求するデータを収集するクライアント管理部と、該クライアントからのレプリカ作成要求を解析するレプリカ作成要求解析部と、解析した作成要求に従ってレプリカを作成し、作成したレプリカの内容と該レプリカを格納するデータベース上の格納位置とに関する情報を対応づけたレプリカ管理テーブルを用いて該レプリカを管理するレプリカ作成管理部とを有し、該レプリカ作成管理部により、前記レプリカ作成要求と、前記データ収集手段で利用可能な計算機資源およびサーバ側で利用可能な計算機資源に基づき、該サーバと交渉を行うことによって、レプリカ作成対象のデータと対価としての料金とを決定するデータ収集側交渉処理と、該サーバで利用可能な計算機資源に基づき、前記データ収集側との間で前記料金を含むレプリカ作成条件に関する交渉を行うサーバ側交渉処理とを実行する、

課金方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、分散ネットワークコンピューティング環境（以下、分散環境）におけるデータウェアハウスシステムとそこで用いられる問合せ処理方法及びそのためのデータ収集方法と装置、さらには、課金システム及び方法に関する。

【0002】

【従来の技術】

計算機システムの低価格化による普及と、高信頼なソフトウェアの出現、社会システムの効率化への要望などから、さまざまな情報がオンライン化され利用されつつある。例えば、企業活動においては、店舗での売り上げ状況や、製品管理情報、顧客情報などの各種業務データが計算機で処理されるようになっている。これらの基幹業務で利用されていた計算機上のデータを、基幹業務のみならず、商品の売れ行きの動向調査や、顧客の嗜好分析などの他の目的で有効活用したいという要望から、データウェアハウスシステムが盛んに用いられるようになっている。データウェアハウスについては、例えば、W. H. Inmon著、"Building the Data Warehouse Second Edition"、John Wiley& Sons, Inc.、ISBN0-471-14161-5に、その構成方法、利用方法が説明されている。データウェアハウスシステムは、その名前の通り、データの倉庫であり、基幹業務で利用される膨大なデータを蓄積・管理するために広まりつつある。。

【0003】

近年、このデータウェアハウスで蓄積・管理されるデータに対して、さまざまな角度から検討を加えることによって、これまで見過ごされていた新しい情報を得られることがわかつってきた。例えば、あるスーパーマーケットで、売り上げデータを解析したところ、「週末に仕事帰りの男性がビールとおむつをまとめ買いする」という一見関係のなさそうな2つの商品間の関連が明らかとなり、この情報をを利用してビールとおむつを近い場所に並べることで売り上げを伸ばすことが

できたという例がある。このように、これまで見過ごされていた有益な情報を、データの中から見つけ出す手法はデータマイニングと呼ばれる。

【0004】

計算機の普及と並行する形で、インターネットに代表されるネットワーク技術の進歩も著しい。このネットワーク技術は、例えば、Robert Orfali、Dan Harkay著、"Client/Server Programming with JAVA and CORBA Second Edition"、John Wiley&Sons、Inc.、ISBN0-471-24578-Xの第1節に説明されている、CORBA (Common Object Request Broker Architecture) に代表される、分散ネットワーク基盤技術とを利用することにより、各種の情報をネットワークを介して利用することが可能となりつつある。

【0005】

以上のような背景から、ネットワーク上に存在する複数のデータベースやデータウェアハウス上のデータを統合し、データマイニングなどの手法を利用することによって、これまでよりもさらに有用な情報を得ようとする試みが生まれるのは自然な成り行きである。データベースの統合利用については、例えばA. Sheth、J. Larson著、"Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases"、ACM Computer Surveys、Vol. 22、No. 3、pp. 183-236や、A. Sheth、G. Karabatis著、"Multi-database Interdependencies in Industry"、Proc. of 1993 ACM Sigmod、Vol. 22、pp. 483-486などに説明されているように、従来から学会を中心として異種データベース、連邦データベース、多データベースなどの研究が盛んであり、複数のデータベースの結合方式が数多く議論されているが、そのほとんどがデータの異種性を考慮して、いかにして相異なるデータを統合利用するかを中心とする研究であった。

【0006】

しかしながら、分散環境でデータウェアハウスシステムを構築しようとする場合には、そのデータ規模が大きいこと、データに対する問合せ処理が従来のデータベース検索処理と比較して複雑なことに起因する性能上の多くの問題が顕れる。例えば、データ規模に関しては、1998年3月現在で数TB（テラバイト；10の12乗バイト）級のデータウェアハウスが構築されている。問合せ処理の複雑さに関しては、データウェアハウス、データマイニングなどの意思決定支援をモデル化し、業界で広く受け入れられている標準的なベンチマーク”TPC-BENCHMARK D (Decision Support) Standard Specification”、Revision 1.2.2、Transaction Processing Performance Councilの問合せがその好適な例となる。例えば、1TBの大規模データに対して、前記TPC-Dの一連の問合せを実行した場合、1998年5月現在の世界最高速コンピュータでも數十分から数時間を要する。

【0007】

データウェアハウスシステムの利用形態としては、図11に示すように、データを蓄積・管理し、記憶装置1105を管理し、問合せ処理を実行するサーバ1102に対して、クライアント1101が処理を依頼し、結果を受け取るというクライアント-サーバ型が一般的である。

【0008】

ところが、図14に示すように、不特定多数の利用者（クライアント）1401～1402が、ネットワーク1405を介して不特定多数のデータウェアハウスやデータベースなどのサーバ1403～1404に問合せ処理を依頼し（1406）その結果を得る（1407）というような、分散環境におけるクライアント-サーバ型の利用形態を想定すると、前述の問合せ処理の負荷の高さから、不特定多数のクライアントからの要求を受け付けるサーバの負荷が高くなり、クライアントからの解析要求に対する処理が遅延することは容易に想像できる。

【0009】

複数のサーバのデータを解析処理の対象とする場合には、クライアント-サーバ

型のデータウェアハウスシステムの拡張として、図12に示すようにサーバの位置情報を管理するモジュール1202がクライアント1201からの問合せ1207を、サーバ位置情報1203を利用してネットワーク1204経由で、サーバ1205～1206に転送し、問合せ処理結果1208がクライアントに返されるという方式が考えられる。例えば、INTERSOLV社のVirtual Data Warehouse System(以下、VDW)がその好適な実現例である。VDWがサーバ位置を管理することにより、クライアント自らは意識することなく、複数のサーバデータをうまく取り扱うことができる。ところがVDWは前記分散環境におけるクライアント～サーバ型のデータウェアハウスシステムと同様に、問合せ処理時のサーバ負荷が高くなってしまい、分散環境におけるデータウェアハウスシステムの好適な実現例とは言いがたい。

【0010】

分散環境において複数のデータベース、もしくはデータウェアハウスに対する問合せを処理するための方式が、特開平8-286960号公報に開示されている。本方式はクライアントの処理負荷を軽減するためにクラスタサーバに問合せを転送する。クラスタサーバは問合せに応じて適切なデータベースに問合せを転送し、該データベースから得られた結果を統合してクライアントに返す。本方式では、問合せは結局サーバに転送されるため、サーバの負荷削減は不可能である。

【0011】

サーバ負荷削減と処理時間短縮という課題に関しては、例えば図13に示すように、サーバ1305～1306のデータ1307～1308をクライアント側のモジュール1309にコピーし(処理1311、1312)、そのコピー1310に対して問合せ1313を発行し、結果1314を得るという方式がある。以下、このサーバデータのコピー1310をレプリカと呼ぶことにする。レプリカに対して問合せ処理を実行することにより、サーバ1305～1306での問合せ処理を避けることができ、サーバ負荷を削減できるとともに、ネットワークを介したサーバへのアクセスを回避できることにより、問合せ処理時間を短縮できる。

【0012】

ところが、分散環境で複数のサーバを対象にしてレプリカを作成しようとする場合には、単純な方法ではクライアント側でレプリカを格納するために大規模な記憶装置1315が必要となる。例えば、300GB（ギガバイト；10の9乗バイト）程度のデータを持つサーバ10台を統合利用しようとするクライアントは単純計算で $300\text{GB} \times 10\text{台} = 3\text{TB}$ もの記憶装置を準備する必要があり、現状の技術ではこの大規模な記憶装置をクライアント側で準備することは不可能に近い。また、レプリカ作成時に大量のデータをネットワーク経由でサーバからクライアントへ転送するため、ネットワークへ大きな負荷がかかってしまう。さらに、レプリカを作成した後で、サーバ側のデータが更新された場合には、対応するレプリカも更新する必要があり、この更新処理コストはレプリカの大きさに比例するため、更新処理のためのコストも無視できなくなってしまい、本方式も分散環境におけるデータウェアハウスシステムの好適な実現例とは言いがたい。

【0013】

これに対して、A. Keller, J. Basu著、"A Predicate-based Caching Scheme for Client-Server Database Architectures", The VLDB Journal, Vol. 5, No. 1, pp. 35-47に記載されているように、レプリカを作成するのではなく、問合せとその問合せ処理結果をキャッシュし、新たな問合せに対しては、キャッシュされた結果を再利用して処理することによって、サーバ負荷の削減、問合せ処理時間の短縮を行う方式が提案されている。本方式は問合せ結果の再利用率が高い場合にはサーバ負荷削減、問合せ処理時間の短縮に大きな効果があるが、分散環境におけるデータウェアハウスを考えた場合には、対象となるデータ規模とクライアント側で準備できる記憶装置の規模の差が極めて大きいため、キャッシュされているデータの再利用率が極めて低くなってしまい、効率が悪いという欠点がある。

【0014】

サーバのファイルをネットワーク経由で取得し、ユーザに提供するための情報処理装置及びシステム及びその制御方式が特開平9-297702号公報に開示

されている。本方式では、システムがユーザからファイル参照要求を受け取った時点でレプリカを作成し、データウェアハウスシステムにおいて、ユーザから問合せが発行された場合に、最初はサーバに検索を依頼する必要が出てくるため、最初の問合せの応答時間の短縮は不可能である。さらに、本方式ではレプリカの作成単位がファイルとなっているため、データベースの問合せ条件に合致したレコードあるいはオブジェクト単位のレプリカ作成は困難である。

【0015】

サーバにおけるデータの更新をクライアント（本願発明における、後述のデータ収集手段に対応）に伝播するための方式に関しては、サーバが例えば1時間毎などの一定の時間間隔で、あるいはサーバデータの更新が起こる毎にクライアントに向けてデータを送信するという、サーバ主導のPUSH方式と、クライアントが一定の時間間隔あるいは必要となった時にサーバにデータを取得しに行くという、クライアント主導のPULL方式がある。ところが、PUSH方式のうち、クライアント個別にサーバがデータを配送する方式ではサーバの負荷が高くなるという問題があり、サーバがデータをブロードキャストあるいはマルチキャストして各クライアントが必要なデータのみを受け取るという方式ではクライアントが適切なタイミングでデータを取得することが困難になるという問題があるので、PUSH方式のみでは分散環境における効率的なデータ配送は困難である。一方、PULL方式では、サーバのデータが更新された場合にはすぐにクライアントのデータも更新する必要があるような制約が厳しい応用を考えた場合にはクライアントは頻繁にサーバのデータをチェックする必要があるが、多数のクライアントが処理要求を頻繁に発行するとサーバにおいてこれらの処理要求を処理するための負荷が高くなってしまうという問題があり、PULL方式のみでも分散環境における効率的なデータ配送は困難であることがわかる。PUSH方式とPULL方式を組合せて利用する方式については、例えばC. Pu, L. Liu著、"Update Monitoring: The CQ Project"、Lecture Notes in Computer Science、Vol. 1368、ISSN 0302-9743、pp. 396-411（以下、CQ）に説明されている。CQでは、クライアントからのトリガ条件を含ん

だ問合せをCQサーバ上に登録し、最初はクライアント主導のPULL方式、次回からは問合せに含まれるトリガ条件に応じてサーバ主導のPUSH方式でデータ転送を行うが、CQでは問合せごとにPUSH方式、PULL方式を指定できるわけではないため、結局データ転送はサーバ主導のPUSH方式となってしまい、サーバ負荷が高くなるというPUSH方式の問題が現れる。

【0016】

【発明が解決しようとする課題】

データベースやデータウェアハウスなどのサーバが多数存在し、不特定多数のクライアントがこれらのサーバにネットワークを介してアクセスし、サーバ上のデータを統合利用して有益な情報を得ようとする場合に、サーバに問合せを転送する方式では、サーバにかかる負荷の増大、ネットワークへの依存度の高さ、問合せへの応答時間の増大が問題となり、クライアント側にレプリカを作成する方式では、大量データ転送に伴うネットワークへの負荷の増大、クライアント側の記憶装置容量の増大、レプリカの更新処理コストの増大が問題となっていた。また、キャッシングを利用する方式では、ヒット率が低くなるため、キャッシングされているデータの再利用率が極めて低くなり、結果として分散環境における効率的なデータウェアハウスシステムの構築は困難であるという問題があった。

【0017】

本発明の目的は分散環境での効率の良いデータウェアハウスシステムとそこで用いられる問合せ処理方法及びそのためのデータ収集方法と装置を提供することである。

【0018】

さらに具体的には、第1の目的はサーバ負荷の削減であり、第2の目的はネットワーク依存度の低減であり、第3の目的は問合せへの応答時間の短縮であり、第4の目的はネットワーク負荷の削減であり、第5の目的はクライアントの記憶容量の削減であり、第6の目的はレプリカの更新処理コストの削減であり、第7の目的はレプリカのヒット率の向上である。

【0019】

【課題を解決するための手段】

上記第1～第3の目的を達成するため、本発明ではクライアントの問合せ処理に用いるサーバデータのレプリカを作成・管理するデータ収集手段を設け、クライアントの問合せ処理を可能な限りデータ収集手段のレプリカを用いて実行する。データ収集手段において問合せ処理を実行することによって、サーバへの問合せ転送が少なくなり、第1の目的であるサーバ負荷が削減できる。また、サーバデータのレプリカをデータ収集手段で作成することにより、サーバへのネットワークが不調となった場合にも該データ収集手段内のレプリカを用いて問合せを処理することが可能となり、第2の目的であるネットワーク依存度が低減できる。さらに、多くのクライアントからアクセスされるために負荷が高くしかも大量のデータを管理するために問合せ処理コストの高い（課金システムでは、料金の高い）サーバにおいて問合せを処理するのではなく、限られた数のクライアントからの問合せを必要なデータのみのレプリカを用いてデータ収集手段において処理することにより、第3の目的である問合せ応答時間が短縮できる。

【0020】

問合せ処理を更に具体的に説明すると、前記データ収集手段はクライアントからの問合せが、自データ収集手段で処理可能か、または協調する他データ収集手段で処理可能か、またはサーバへの転送が必要かを判断し、自データ収集手段、または協調する他データ収集手段で処理が可能な場合には、該当のデータ収集手段で問合せを処理する。協調データ収集手段を用いることにより、データ収集手段で処理可能な問合せが増加し、前記第1～第3の目的達成に寄与する。さらに、前記サーバへの前記問合せの転送が必要だと判断された場合にも、該問合せのうち自データ収集手段および協調する他データ収集手段で利用可能なレプリカで処理しきれない部分だけをサーバに転送することにより、サーバで処理するデータ量が少なくなり、サーバ負荷が削減できるだけでなく、サーバからクライアントに返される結果データも少なくなるため、第4の目的であるネットワーク負荷が削減できる。

【0021】

さらに、単純にレプリカを作成した場合に生じる前記課題を解決して前記第4～第6の目的を達成するため、データの共有が可能なクライアントをグループ化し、グループ化したクライアント集合に対してレプリカを作成し、該レプリカをクライアント間で共有すると共に、必要に応じて自データ収集手段は、協調する他データ収集手段と連係する。レプリカを共有することにより、サーバからデータ収集手段に転送されるデータ量が削減できるため、ネットワーク負荷が削減できる。本発明ではデータ収集手段においてレプリカを作成するため、第5の目的であるクライアントが必要とする記憶容量を削減できることは無論のこと、データ収集手段間でのレプリカ共有により、データ収集手段における記憶容量も削減できる。また、レプリカの共有により、システム全体としてのレプリカの量が削減できるため、第6の目的であるレプリカの更新処理コストも削減できる。

【0022】

レプリカ作成に際しては、ユーザからクライアントを介して与えられるデータの精度、鮮度、および優先度などのデータの質に関する条件と、収集するデータの範囲に関する条件とを含むレプリカ作成要求を受け付けて該要求を保持し、記憶装置容量、CPU性能などの該データ収集手段が利用可能な資源量を考慮しながら、データを供給するサーバと交渉処理を行うことによって、前記レプリカ作成要求の一部あるいは全部を満足するレプリカを作成する。レプリカ作成要求を、ユーザから与えてもらうことにより、ユーザの意図するデータの収集が可能となり、第7の目的であるレプリカのヒット率を向上できる。また、レプリカを作成する際にデータの質を調整することにより、データ収集手段の計算機資源に応じた適切なサイズのレプリカを作成でき、第4の目的であるネットワーク負荷の削減、第5の目的であるクライアント及びデータ収集手段の記憶装置容量の削減、第6の目的であるレプリカの更新処理コストの削減ができる。

【0023】

レプリカの更新処理においては、クライアント主導のPULL手法とサーバ主導のPUSH手法を組み合わせて用いることにより、クライアントのデータに対する要求を考慮しながら、サーバの負荷を削減でき、レプリカ更新時のサーバ負

荷を削減できる。これにより第1の目的が達成できる。

【0024】

【発明の実施の形態】

図1に、本発明によるデータウェアハウスシステムの好適な実現例を示す。クライアント103、104は内部ネットワーク128を介してデータ収集手段1(101)に接続される。

【0025】

内部ネットワーク128は、イーサネット、光ファイバ、FDDIで接続されるローカルエリアネットワークであってよく、クライアントはHitachi FLORAなどのパーソナルコンピュータ、Hitachi 3050クリエイティブワークステーションなどの任意のコンピュータ・システムでよい。データ収集手段1は複数のクライアントをグループ化して管理し、クライアントから発行されるレプリカ作成要求、および問合せを受け付け、レプリカ作成要求をレプリカ作成要求解析部106へ、問合せを問合せ解析部109に転送するクライアント管理部105と、クライアントからのレプリカ作成要求に対して実際にレプリカを作成するか否かを決定し、レプリカを作成する場合には、作成するレプリカに関する情報であるレプリカ記述をレプリカ作成管理部107に転送するレプリカ作成要求解析部106と、レプリカ管理テーブル108を参照しながらレプリカ123を記憶装置112に格納・管理するレプリカ作成管理部107と、クライアントからの問合せを解析する問合せ解析部109と、クライアントからの問合せがデータ収集手段1で処理できる場合に該問合せを処理する問合せ処理部127と、ネットワーク113を介してサーバや他のデータ収集手段との通信処理を管理する通信制御部110とを含む。さらに高度な処理を行わせるため、データ収集手段にはデータ収集手段交渉処理部111を備えるようにしても良い。該データ収集手段交渉処理部については後述する。データ収集手段は、クライアントと同様に任意のコンピュータ・システムでよく、記憶装置112は磁気記憶装置、光ディスク装置、磁気テープなどでよい。ネットワーク113は前記ローカルエリアネットワークでも、地理的に分散された複数のサイトを接続する広域エリアネットワークでもよい。

【0026】

サーバ1(114)はデータ収集手段からのレプリカ作成要求や問合せを受け付ける通信制御部115と、問合せを処理する問合せ処理部117と、データ収集手段へ配達するデータを管理する配達データ管理部118と、配達データ管理部が参照する配達データ管理テーブル120とを含む。さらに高度な処理を行わせるため、サーバ交渉処理部116および負荷管理部119を備えるようにしても良いが、これらについては後述する。サーバもデータ収集手段と同様に任意のコンピュータ・システムでよく、サーバデータ124を格納する記憶装置121は磁気記憶装置、光ディスク装置、磁気テープ、CD-ROMなどの任意の記憶装置、あるいはこれらの記憶装置の組み合わせでよい。さらに、サーバデータおよびデータ収集手段のレプリカデータの管理はファイルシステム、あるいはHITACHI HiRDBなどの汎用のデータベースマネージメントシステムで行って差し支えない。

【0027】

なお、前記データ収集手段内のクライアント管理部、前記レプリカ作成要求解析部、前記レプリカ作成管理部、前記問合せ処理部、前記通信制御部、及び前記データ収集手段交渉処理部、ならびに前記サーバ上の通信制御部、前記問合せ処理部、前記サーバ交渉処理部、前記負荷管理部、及び前記配達データ管理部は、専用のハードウェアとして構成されるばかりでなく、それぞれデータ収集手段内の記憶装置112、ならびに前記サーバ内の記憶装置121にローカルに格納されたプログラムであったり、あるいはネットワーク上のプログラムを格納するサーバからダウンロードしたプログラムであってよい。異種分散環境において、サーバからダウンロードしたプログラムを安全に実行するしくみについては、例えば、Mary Campione、Kathy Walrath著、"The Java Tutorial"、Addison-Wesley、ISBN 0-201-63454-6の4節で説明されているようなプログラミング言語（以下、インターネットプログラミング言語）を利用できる。

【0028】

特に、クライアントの好適な実施例としてはHitachi FLORAなどの

パーソナルコンピュータ、Hitachi 3050クリエイティブワークステーションなど任意のコンピュータシステム上で、マイクロソフト社のインターネットエクスプローラ、ネットスケープ社のネットスケープナビゲータなどのウェブ・ブラウザを用いる形態が考えられるが、この場合には前記インターネットプログラミング言語を用いて作成したプログラムモジュールをブラウザにダウンロードすることにより、クライアントプログラムの動的な変更が可能となる。

【0029】

なお、本実施例ではクライアントが内部ネットワークを介して直接的にデータ収集手段1～nに接続されている例を示したが、図18に示すようにクライアントがLANやインターネットを介してデータ収集手段に接続されている場合においても本発明が有効であることはいうまでもない。

【0030】

本発明のさらに具体的な特徴は、（1）データ収集手段内にグループ化したクライアントで共有が可能な部分的レプリカを作成する。（2）レプリカの作成にはユーザからのレプリカ作成要求を受け付け、サーバとの交渉処理によって、実際にレプリカを作成するデータを決定し、その内容を記述したレプリカ記述を作成・管理する。（3）レプリカ作成要求にはデータの範囲に関する条件（データ領域条件）だけではなく、データの質に関する条件（データ品質条件）を含む。

（4）レプリカ更新には、サーバ側で保持する配達データ管理テーブルを利用して、サーバ主導のPUSH手法、クライアント主導のPULL手法を組合せたデータ転送方式を用いる。（5）自データ収集手段内のレプリカを用いて処理可能な問合せは該自データ収集手段のレプリカを用いて処理し、逆に、不可能な問合せは、処理が可能な他データ収集手段、もしくはサーバに転送し問合せを依頼する、という5ポイントにまとめることができる。以下それについて、例を用いながら説明する。

【0031】

（1）従来技術では、図14で示すような単純なクライアント-サーバ形態では、分散環境におけるデータウェアハウスシステムの実現は困難であると述べた。そこで、本発明ではデータの共用が可能なクライアントをまとめて管理するデー

タ収集手段を設ける。このデータ収集手段を利用することにより、クライアント側に重複する冗長なレプリカの作成を抑制でき、無駄なサーバへのアクセスを防止できるので、サーバへかかる負荷を軽減できる。例えば、図15において、クライアント1(1501)は売上げが10000円以上の商品売上表のデータ(1503)を必要とし、クライアント2(1502)は売上げが5000円以上の前記商品売上表のデータ(1504)が必要とする。商品売上げデータ1509はサーバ1505に格納されているとすると、クライアント1、2はそれぞれ該サーバからデータを取得する必要がある。つまり、サーバにはクライアント1と2の両方からの問合せによる負荷がかかる。ところが、本発明のデータ収集手段1506を設置し、該データ収集手段がクライアント1および2のレプリカ作成要求を保持し、その条件の和集合である、「商品売上げが5000円以上のデータ」(1507)をレプリカ記述として採用してそのデータをサーバから取得し、レプリカ1508を作成すれば、クライアント1、2の問合せはデータ集手段のデータを用いて処理できるため、クライアント1、2のサーバへのアクセスを省くことができ、サーバ負荷およびデータ転送によるネットワーク負荷を削減することができる。また、データ収集手段の記憶装置1610の容量の制約、あるいはデータ収集手段1606のCPU処理能力の制約、あるいはネットワークを転送するデータ量の制約から、クライアント1および2が要求するレプリカ作成要求の和集合が保持できない場合には、図16に示すように、クライアント1(1601)とクライアント2(1602)のそれぞれのレプリカ作成要求(1603、1604)から、その条件の積集合である「商品売上げが10000円以上のデータ」(1607)をレプリカ記述としてそのデータをサーバから取得し、レプリカを作成すればクライアント1からの処理要求がレプリカ作成要求を満たす場合には、全ての処理要求をレプリカを用いて処理でき、クライアント2についてもレプリカを用いて処理できない処理要求つまり「商品売り上げが5000円以上10000円未満のデータ」を参照する問合せのみをサーバに転送すればよく、サーバ負荷およびネットワーク負荷を削減することができる。図15、16を用いて説明したサーバ負荷削減、ネットワーク負荷削減の効果は、クライアント数が多くなるほど大きくなることは明らかである。

【0032】

(2) レプリカ作成要求の受け付けと、レプリカ作成方法について図1および図2および図8を用いて説明する。データ収集手段101がクライアント103～104からのレプリカ作成要求（レプリカ作成要求に関しては（3）で説明する）を受けつけた場合（処理202）には、レプリカ作成要求解析部106がその要求を解析し（処理203）、レプリカ作成管理部107と通信することにより、内容を図8に示すレプリカ管理テーブル108を参照する（処理204）。レプリカ管理テーブルは、作成されるデータの領域に関するデータ領域条件801と、データの品質に関するデータ品質条件802のエントリを含む。この2つのエントリをレプリカ記述803と呼ぶ。一つの前記レプリカ記述のエントリに対し、該レプリカが格納されているデータ収集手段名を記述するレプリカ位置情報804と、該レプリカのデータ源であるサーバ名を記述するサーバ位置情報805と、該レプリカのメンテナンス条件であるデータ配送条件806とで構成される。例えば図8の1番目のエントリ807は、注文表に格納されているレコードのうち、価格が10000円以上のレコードの、注文番号と、価格と、顧客番号の3つのカラムを取出したデータのレプリカがデータ収集手段1にあり、該レプリカの元データはサーバ1であり、サーバ1からは13:00にPUSH手法でデータが配送され、レプリカがメンテナンスされることを表している。但し、図8の2番目のエントリにおけるデータ配送方法で、{1:00, 13:00}、PUSHという記述は、1:00と13:00の両方の時刻にPUSH手法でデータが配送されることを表す。

【0033】

要求されたレプリカが自データ収集手段内の既存のレプリカから作成可能な場合（判定処理205でYesが選択された場合）には、新たなレプリカ作成を行わずにレプリカ作成処理を終了する。例えば、レプリカ管理テーブルが図8に示されるような場合で、新たなレプリカ作成要求が図19の1番目のエントリ1901であるとすると、該レプリカ作成要求は図8の807に示される、データ収集手段1内にある既存のレプリカを用いれば処理でき、判定処理205はYesとなる。

【0034】

要求されたレプリカが自データ収集手段内の既存のレプリカ123から作成不可能な場合（判定処理205でNoが選択された場合）には、要求するレプリカが協調するデータ収集手段102の既存のレプリカ126から作成可能かどうかを判定する（判定処理208）。もし作成可能な場合（判定処理208でYesが選択された場合）には、自データ収集手段内にさらにレプリカを作成するかどうかを判定する（判定処理216）。自データ収集手段内には重複してレプリカを作成しない場合（判定処理216でNoが選択された場合）には、レプリカ作成を行わずにレプリカ作成処理を終了する。自データ収集手段内に重複してレプリカを作成する場合（判定処理216でYesが選択された場合）には、クライアントから要求された条件で、レプリカを協調データ収集手段102のレプリカから作成し（処理215）、レプリカ作成処理を終了する（219）。

【0035】

例えば、レプリカ管理テーブルが図8に示すような場合で、新たなレプリカ作成要求が、図19の2番目のエントリ1902であるとすると、データ収集手段1内の既存のレプリカでは新たなレプリカ作成要求を処理できないが、図8の808に示される、データ収集手段2内のレプリカを用いれば、該新レプリカ作成要求を処理できる。このような場合、データ収集手段1の記憶装置に該レプリカを作成する余裕がある場合には、データ収集手段2内のレプリカからデータ収集手段1内に新たにレプリカを作成する。また、データ収集手段1に対するクライアントからの該レプリカ作成要求の優先度が高い場合には、データ収集手段1内の優先度の低いレプリカを消去して該レプリカ作成要求に従うレプリカを作成する。

【0036】

既存のレプリカから新たなレプリカ作成要求に基づくレプリカが作成できるかどうかは、既存のレプリカの内容を記述するレプリカ管理テーブルと、新たに与えられたレプリカ作成要求を比較することによって判断する。つまり、データ収集手段はレプリカ作成要求を保持し、該レプリカ作成要求の集合と、該レプリカ作成要求の組合せから生成されているレプリカ記述に対して、新たなレプリカ作成

要求が届いた場合には、該データ収集手段は該レプリカ作成要求を保持し、該レプリカ作成要求と前記レプリカ管理テーブルのエントリとを比較し、同値関係、包含関係を決定する。例えば、先の例では図8に示されるレプリカ管理テーブルの各エントリに対して、図19で示される新たなレプリカ作成要求が与えられた場合、エントリ1902で表される新レプリカ作成要求は、エントリ808で表される既存のレプリカに包含されるのは自明である。なお、本実施例では、レプリカ作成要求が非常に単純な場合を想定したが、さらに一般的な場合についても、例えば、Jeffrey D. Ullman著、"PRINCIPLES OF DATABASE AND KNOWLEDGE-BASE SYSTEMS"、VOLUME II、COMPUTER SCIENCE PRESS、ISBN 0-7167-8162-X、第14章"Optimization for Conjunctive Queries"は、"Query Equivalence、Query Containment"と呼ばれる方式を用いることにより、条件間の同値関係、包含関係を調べる方法が開示されている。該判定方式を利用することにより、さらに一般的な場合についても、本発明によるレプリカ管理テーブルを用いたレプリカ管理が実現できる。

【0037】

クライアントから要求されたレプリカが自データ収集手段内の既存レプリカからも、協調データ収集手段内の既存のレプリカからも作成できない場合（判定処理208でNoが選択された場合）には、サーバとの交渉処理を行う（処理211）。この交渉処理の結果、データ収集手段およびサーバの双方が承諾できるレプリカ作成条件が存在しなかった場合（判定処理212でNoが選択された場合）には、レプリカ作成を行わずにレプリカ作成処理を終了する（219）。交渉処理の結果、双方が承諾できるレプリカ作成条件が存在した場合（判定処理212でYesが選択された場合）には、双方が承諾したレプリカ作成条件に従ってレプリカを作成し（処理215）、レプリカ作成処理を終了する（219）。

【0038】

ステップ211の交渉処理について図1および図3を用いて説明する。まずレプリカ作成要求がデータ収集手段101からサーバ114に転送される（処理30

2) と、そのレプリカ作成要求をサーバが受け入れる場合（判定処理303でY_esが選択された場合）には、レプリカ作成条件（課金システムでは、料金）をデータ収集手段が要求したレプリカ作成条件とし（処理310）、作成された条件に基づいてデータ収集手段のレプリカ作成管理部107がレプリカ管理テーブル108と、サーバの配送データ管理部118が配送データ管理テーブル120を更新して、交渉処理を終了する（311）。データ収集手段から転送されたレプリカ作成要求をサーバが受け入れられない場合（判定処理303でN_oが選択された場合）で、サーバからデータ収集手段に対して提示できる新しい条件が存在しない場合（判定処理312でN_oが選択された場合）には、レプリカ作成要求に関する情報を設定することなく交渉処理を終了する（311）。データ収集手段から転送されたレプリカ作成要求をサーバが受け入れられない場合（判定処理303でN_oが選択された場合）で、サーバからデータ収集手段に対して提示できる新しい条件が存在する場合（判定処理312でY_esが選択された場合）には、サーバは新しい条件をデータ収集手段に転送する（処理306）。データ収集手段が、サーバが提示した条件を受け入れる場合（判定処理307でY_esが選択された場合）には、新しい条件をレプリカ作成条件として（処理310）、データ収集手段側のレプリカ管理テーブルおよびサーバ側の配送データ管理テーブルを更新して、交渉処理を終了する（311）。データ収集手段が、サーバが提示した条件を受け入れられない場合（判定処理307でN_oが選択された場合）には、レプリカ作成要求に関する情報を設定することなく交渉処理を終了する。

【0039】

交渉処理の具体例を図1および図10を用いて説明する。クライアント103～104からのレプリカ作成要求が図10（A）に示すような条件であったとする。まずデータ収集手段1（101）はレプリカ作成要求1001をネットワーク113を介してサーバ1（114）に転送する。サーバ1は負荷管理部119で管理される現在の自システムの負荷から、該レプリカ作成要求を受け付けた場合の自システムの負荷を予測し、その予測値がある閾値以下の場合には該レプリカ作成要求を受け付け、図10（B）に示すように、前記データ収集手段に受け付

けという応答（1003）を返し、該レプリカ作成要求がレプリカ作成条件となって交渉処理は終了する。これに対して、レプリカ作成要求1002がサーバm（122）に転送された場合で、前記負荷予測の結果、サーバmが作成要求1002に示される条件でのレプリカ作成要求を受け付けられないが、作成要求1004に示されるような新条件であれば受け付け得るという場合には、サーバmはデータ収集手段1に条件付受け付けという応答と、サーバで生成した新条件を返す。データ収集手段1は該新条件をレプリカ作成要求を発行したクライアント103あるいは104に返し、クライアントが該新条件を受け付け得る場合には、該新条件をレプリカ作成条件としてデータ収集手段側のレプリカ管理テーブルおよびサーバ側の配送データ管理テーブルを更新して、交渉処理を終了する。

【0040】

自システムの負荷計測と、前記レプリカ作成要求を受け付けた場合の負荷予測については、以下のような実施例が考えられる。一般的にマルチタスクのオペレーティングシステムは”run queue”と呼ばれる実行可能なプロセスの待ち行列を持ち、この行列の平均待ち行列長の時間平均をロードアベレージと呼ぶ。そこで、例えばサーバは自システムのロードアベレージLを自システムの負荷とみなし、前記新レプリカ作成要求を受け付けた場合の自システムの負荷をL+1と見積もり、該予測負荷がある閾値 L_{MAX} 以下の場合（つまり $L+1 \leq L_{MAX}$ ）には前記レプリカ作成要求を受け付ける。また、ロードアベレージを計測不可能なサーバの場合には、一定時間あたりの平均ジョブ数Nを自システムの負荷とみなし、前記新レプリカ作成要求を受け付けた場合の自システムの負荷をN+1と見積もり、該予測値がある閾値 N_{MAX} 以下の場合（つまり $N+1 \leq N_{MAX}$ ）には前記レプリカ作成要求を受け付けるという実施例が考えられる。

【0041】

（3）まず、レプリカ作成要求について説明する。例えば、図15に示す実施例においては、クライアント1、2は商品の売上げデータを収集し、解析を行う。データマイニングにおいては、例えばあるクライアントでは東京都のみの売上げデータを解析対象とするというように、解析対象となるデータは何らかの制約に基づいている場合が多い。この制約は、クライアントを用いて解析を行おうとす

る人間の意図に基づいており、この意図を計算機が自動的にくみとるのは現状では困難であり、無駄も多い。そこで本発明では、クライアントに、アクセスするデータに関するレプリカ作成要求をユーザがシステムに与えられるようなインターフェースを設ける。該インターフェースの実装は、データ収集手段、サーバの場合と同様に、図15のクライアント1501、1502上にローカルに格納されたプログラムであるか、あるいはネットワーク上のプログラムを格納するサーバからダウンロードしたプログラムであってよい。該クライアントの該インターフェースを利用してユーザから発行されたレプリカ作成要求に従って、データ収集手段1506にクライアントが共用可能なレプリカ1508が作成される。

【0042】

前記レプリカ作成要求は、図6に示すように、データの範囲を示すデータ領域条件(601)と、データの質を示すデータ品質条件(603)と、データの配送方法を示すデータ配送条件(604)とを含む。データの質に関する条件には、例えばサーバで更新済みのデータも、レプリカでは更新前のデータを1時間以内は最新データであるとみなすというような鮮度条件605、注文表の注文番号に対する10%のサンプルをレプリカ対象データとする精度条件606、注文表の価格の上位10傑のデータを対象とするというような優先度条件607を含む。但し、従来の問合せ言語との互換性を考慮し、該レプリカ作成要求はデータの質に関する条件を含まない記法も許すこととする。

【0043】

レプリカ作成要求が鮮度条件を含まないときにクライアントが最新のデータを要求した場合には、必ずサーバのデータをチェックする必要があることとし、レプリカ作成要求が精度条件を含まない場合は、精度は100%とすることとし、優先度条件を含まない場合は全ての順位に対するデータを求めるこことする。精度条件については、図7に示すように、データのタイプに応じて、関係データベースレコードに対するサンプリング、フィールド切り出し、文書に対する要約作成、キーワード切り出し、静止画像に対する可逆、非可逆圧縮、輪郭抽出、色数削減、解像度削減、サイズ縮小、動画像に対するフレーム数削減、フレーム内画像圧縮、音声に対する音質調整、文字データへの変換などの各種方法の適用形態が

考えられる。

【0044】

データ品質条件をレプリカ作成要求に含むことによって、レプリカのサイズを小さくすることができる。精度の調整に関しては、特開平09-025863号公報「データベース処理システムにおける集計結果推定方式」に開示されている方式を用いることにより、小規模のサンプリングデータで高精度の結果推定が行えるため、データウェアハウスシステム構築には非常に有効であることがわかる。また、鮮度条件を用いることにより、分散システムで議論の多い更新の問題を柔軟に取り扱うことが可能となる。

【0045】

例えば、図17に示すように、鮮度に関する制約が1日と指定されている場合で、レプリカの前回の更新時刻が1997年10月23日の6:00であり、現在の時刻が同年10月24日の22:00であるとする。実際のサーバ上のデータはこの間も更新されている可能性があるが、鮮度に関する制約が1日として与えられている場合には、現時刻のデータ D_A (T_C) を前回更新時刻のデータ D_A (T_P) とみなすことにより、サーバのデータが更新される毎にネットワークを介したデータ転送が行われることを防ぎ、ネットワーク負荷削減が可能となる。もちろん、サーバデータが更新された時にすぐに更新が反映される必要があるような利用形態に対しては、鮮度条件を指定しないようにすればよい。

【0046】

さらに、図6に示すような優先度に関する制約607を適用するとデータの転送量が削減できる。この制約607は注文表の価格の上位10傑をレプリカの対象とするという指定であり、注文表全体のレプリカを作成する場合と比較して転送データ量は遥かに少なくなる。

【0047】

(4) レプリカ更新時のデータ配送条件について、図1および図5を用いて説明する。サーバ1(114)の管理するデータ124が更新された場合(処理502)には、サーバの配送データ管理部118は配送データ管理テーブル120を参照し(処理503)、更新されたデータが配送データ管理テーブルに登録され

ていない場合（判定処理504でNoが選択された場合）には、データの配達を行うことなく更新処理を終了する（506）。更新されたデータが配達データ管理テーブルに登録されている場合（判定処理504でYesが選択された場合）には、配達管理テーブルのデータ配達条件に従って、データ配達先に該当するデータを配達し（処理505）、更新処理を終了する。

【0048】

レプリカ更新処理の具体例を、図1および図9を用いて説明する。サーバ1（114）は記憶装置121に注文表データ124を格納している。該注文表に新しい注文データ907が挿入されたとする。配達データ管理部118は、該更新データを配達データ管理テーブルと照合し、該更新データが配達データ管理テーブルのエントリ905、および906を満たしているため、該更新データをエントリ905にしたがって、データ収集手段1に13:00にサーバ主導のPUSH手法で、エントリ906にしたがって、データ収集手段2に1:00と13:00にサーバ主導のPUSH手法で配達する。13:00のデータ配達の際には、サーバ1はマルチキャストを行うことにより、データ収集手段1とデータ収集手段2のためのデータ配達を一括して行うことができるため、サーバ負荷およびネットワーク負荷が削減できる。さらに進んで、もしもエントリ906のデータ配達条件が11:00から15:00の間に1度とか、一日に1度など、エントリ905のデータ配達条件を包含するような場合には、エントリ906のデータ転送をエントリ905のデータ転送条件に合わせることにより、データ転送回数を削減することができる。

【0049】

（5）レプリカを用いた問合せ処理について、図1および図4を用いて説明する。クライアント103からデータ収集手段101に問合せが発行されると、問合せはクライアント管理部105を介して、問合せ解析部109に転送される。問合せ解析部はまずレプリカ作成管理部107と通信してレプリカ管理テーブルを参照する（処理403）。前記問合せ解析部は、レプリカ管理テーブルから、問合せが自データ収集手段内で処理可能かどうかを判定し、その問合せが自データ収集手段内で処理可能な場合（判定処理404でYesが選択された場合）に

は、自データ収集手段内で問合せ処理を行い（処理407）、クライアントに解を転送して（処理415）、問合せ処理を終了する（416）。自データ収集手段内での処理が不可能な場合（判定処理404でN○が選択された場合）には、協調する他データ収集手段で問合せが処理できるかどうかを判定し、処理が可能な場合（判定処理408でY e sが選択された場合）には、協調する他データ収集手段に問合せを転送し（処理411）、解を受け取り（処理412）その解をクライアントに転送し（415）、問合せ処理を終了する。自データ収集手段内でも協調する他データ収集手段でも問合せが処理できない場合（判定処理409でN○が選択された場合）には、サーバへ問合せを転送し（処理413）、解を受け取り（処理414）、クライアントに転送し（処理415）、問合せ処理を終了する（416）。

【0050】

クライアントから与えられた問合せが、レプリカを用いて処理できるか否かは、既存のレプリカの内容を記述するレプリカ管理テーブルと、与えられた問合せを比較することによって判定する。つまり、該レプリカ管理テーブルのエントリ（以下、エントリ）と該問合せとを比較し、該問合せが該エントリと同値、あるいは該問合せが該エントリに包含される場合には該レプリカを用いて該問合せが処理できる。本判定は、既存のレプリカを用いて前記新レプリカ作成要求が処理できるかどうかと同値であり、前記新レプリカ作成要求の処理で説明したように、一般的な場合についても、例えばJeffrey D. Ullman著、"PRINCIPLES OF DATABASE AND KNOWLEDGE-BASE SYSTEMS"、VOLUME II、COMPUTER SCIENCE PRESS、ISBN 0-7167-8162-X、第14章"Optimization for Conjunctive Queries"に開示されている、"Query Equivalence、Query Containment"と呼ばれる方式を用いることにより、前記問合せを前記レプリカを用いて処理できるかどうかを判定できるため、本発明による自データ収集システムのレプリカ、協調する他データ収集システムを用いた問合せ処理が実現できる。

【0051】

レプリカを用いた問合せ処理の具体的な例を、図1および図8を用いて説明する。クライアント103から「価格が20000円以上の注文の注文番号と、その注文の価格を求めよ」という問合せが発行されたとする。クライアントからの問合せを受け取ったデータ収集手段1のクライアント管理部105は問合せを問合せ解析部101に転送する。該問合せ解析部101では、レプリカ管理テーブル108を参照し、与えられた問合せはエントリ807に対応する、自データ収集手段1内のレプリカで処理できることを判定し、該レプリカを用いて問合せ処理を行う。これにより、サーバに対して問合せ処理が転送されることはない。次に、「価格が3000円以下の注文の注文番号を求めよ」という問合せがクライアント103から発行された場合を考える。この場合には、レプリカ管理テーブルを参照することによって、該問合せがエントリ808に対応する、データ収集手段2内のレプリカで処理できることを判定し、問合せをデータ収集手段2に転送する。最後に、「価格が7500円の注文をした顧客番号を求めよ」という問合せが発行された場合を考える。この場合に、図8のレプリカ管理テーブルでは該問合せを処理するためのエントリは存在せず、該問合せはレプリカを用いての処理は不可能と判定され、問合せはサーバに転送される。

【0052】

以上の説明において、データ収集手段側では、前記CORBA等の分散ネットワーク基盤技術を利用することにより、必要なデータを取得できるサーバ名を得られると仮定しているが、もしそれが不可能な場合にも、データ収集手段側でサーバに関する情報を管理することによって、本発明を実施できることはいうまでもない。

【0053】

【発明の効果】

本発明のデータウェアハウス構成方法に基づく、例えば、クライアントのグループ化により、クライアント間でのデータ共有が可能となり、クライアント側の記憶装置容量、更新処理コスト、ネットワーク負荷を削減できる。また、レプリカの作成の際にユーザから与えられたデータの要求条件に基づく、データ供給元

のサーバとの交渉処理によって、レプリカを作成するデータ収集手段の計算機資源を考慮しながら、ユーザからの問合せに対して利用率が高いレプリカを作成できる。前記レプリカを用いてクライアントからの問合せを処理することによってサーバ負荷を削減でき、実用的なデータウェアハウスシステムの構築とこれを用いた問合せ処理方法が実現可能となる。

【図面の簡単な説明】

【図1】

本発明によるデータウェアハウスシステムの第1の実施例のブロック構成図。

【図2】

レプリカ作成処理手順を示すフローチャート。

【図3】

交渉処理手順を示すフローチャート。

【図4】

問合せ処理手順を示すフローチャート。

【図5】

更新処理手順を示すフローチャート。

【図6】

本発明の実施例における、クライアントから与えられるレプリカ作成要求を示す図。

【図7】

本発明の実施例における、データタイプに応じた精度調整方法を示す図。

【図8】

本発明の実施例における、レプリカ管理テーブルを示す図。

【図9】

本発明の実施例における、配送データ情報、および更新データを示す図。

【図10】

本発明の実施例における、交渉処理時に利用される情報を示す図。

【図11】

従来技術における、データウェアハウスの利用法を示す図。

【図12】

従来技術における、分散環境でのデータウェアハウスの構成方法を示す図。

【図13】

従来技術における、分散環境でのデータウェアハウスの構成方法を示す図。

【図14】

従来技術における、分散環境でのクライアントとサーバの多対多の構成方法を示す図。

【図15】

本発明によるデータ収集手段でのクライアントのグループ化方法の1例を示す図。

【図16】

本発明によるデータ収集手段でのクライアントのグループ化方法の他の例を示す図。

【図17】

本発明の実施例における、鮮度条件の利用方法を示す図。

【図18】

本発明によるデータウェアハウスシステムの第2の実施例のブロック構成図。

【図19】

本発明におけるレプリカ作成要求の1例を示す図。

【符号の説明】

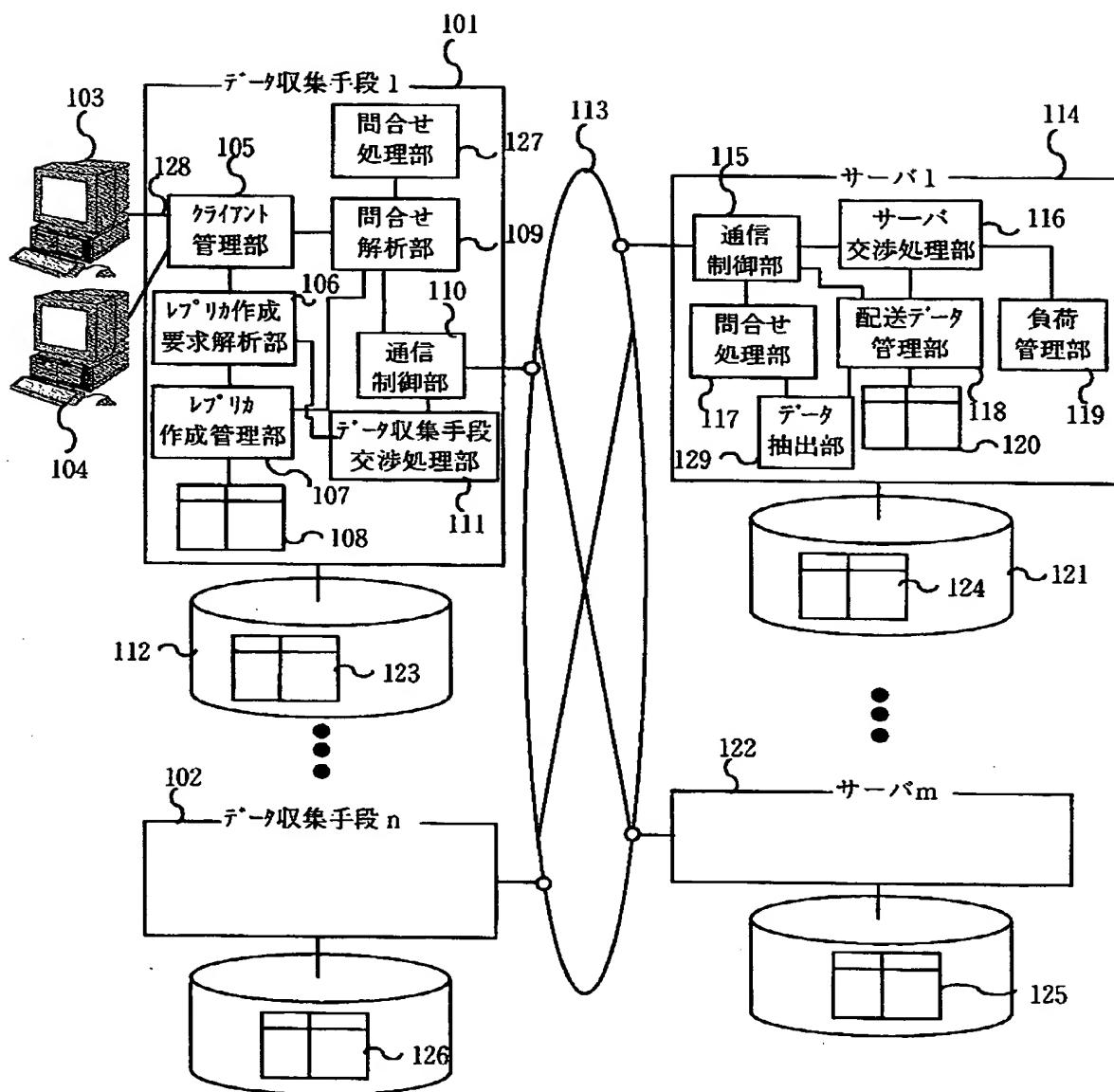
103、104、1101、1201、1301、1401、1402、1501、1502、1601、1602、1806、1807…クライアント端末、
112、121、1105、1209、1210、1315、1316、1317、1408、1409、1510、1511、1610…記憶装置、113、
1204、1304、1405、1802…ネットワーク、114、122、1102、1205、1206、1305、1306、1403、1404、1505、1605、1803、1805…サーバ、101、102、1506、1606、1801、1804…データ収集手段、1202…サーバ位置情報管理モジュール、1203…サーバ位置情報、123、126、1310、1508

、1608…レプリカ、124、125、1307、1308、1509、1609…サーバの管理するデータ、108…レプリカ管理テーブル、120…配達データ管理テーブル、1507、1607…レプリカ記述。

【書類名】 図面

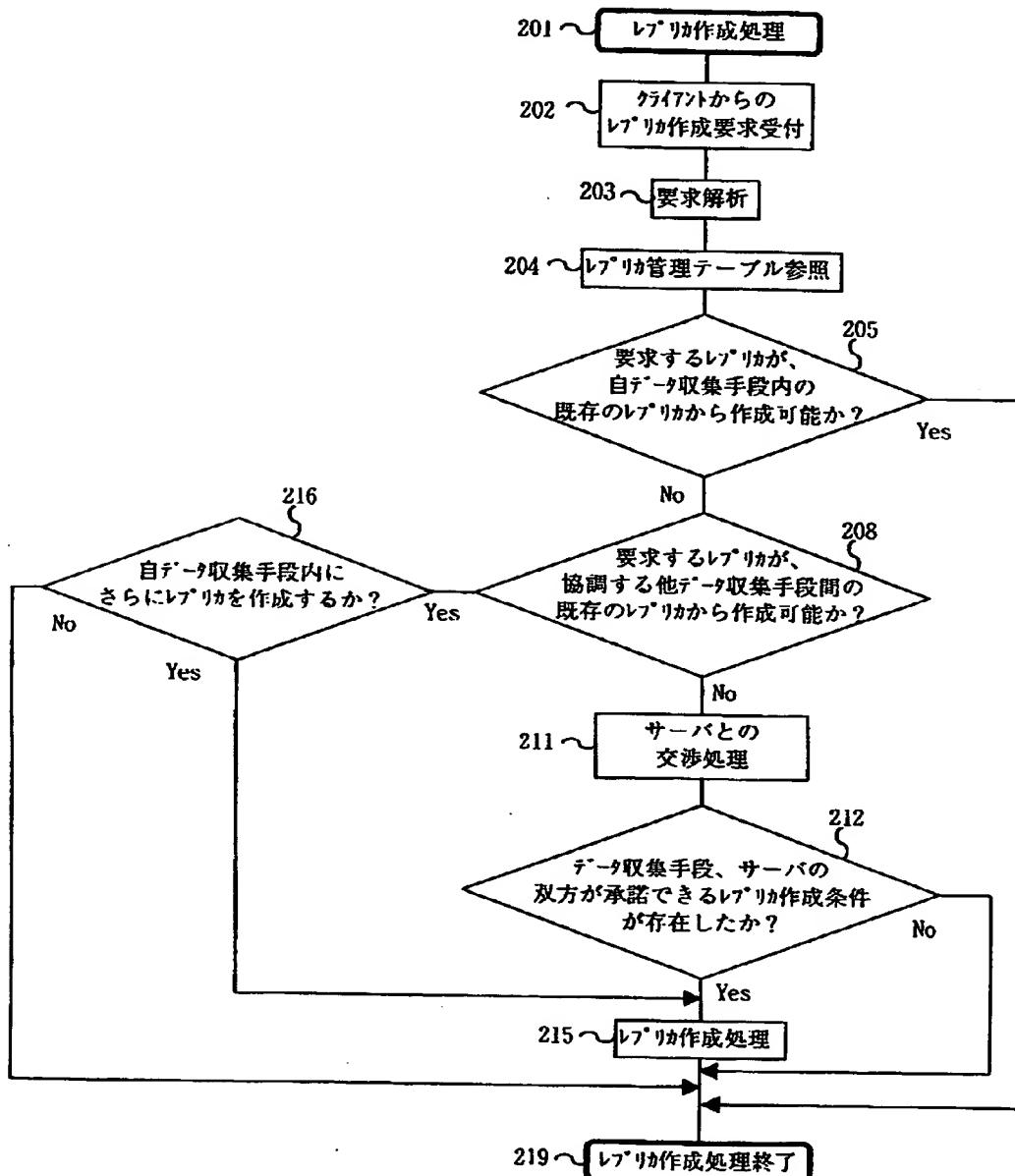
【図1】

図1



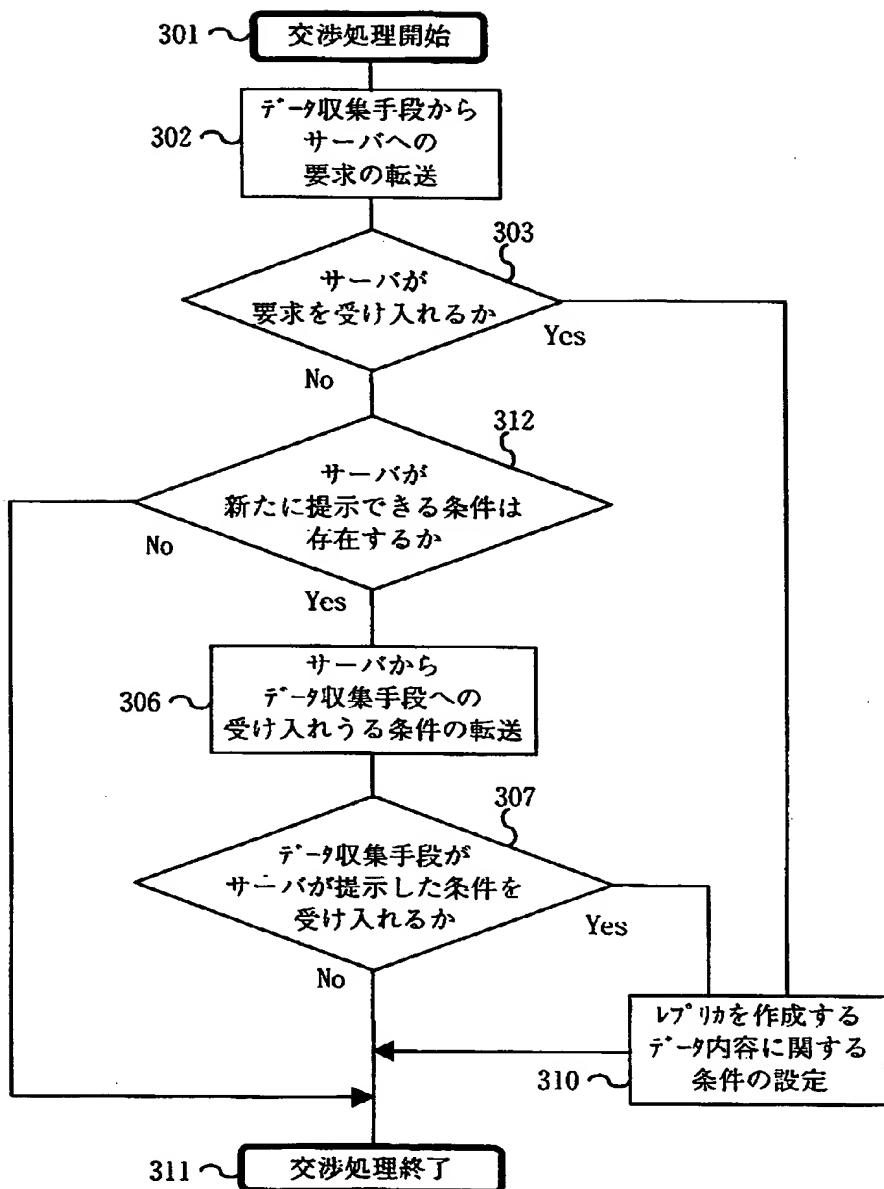
【図2】

図2



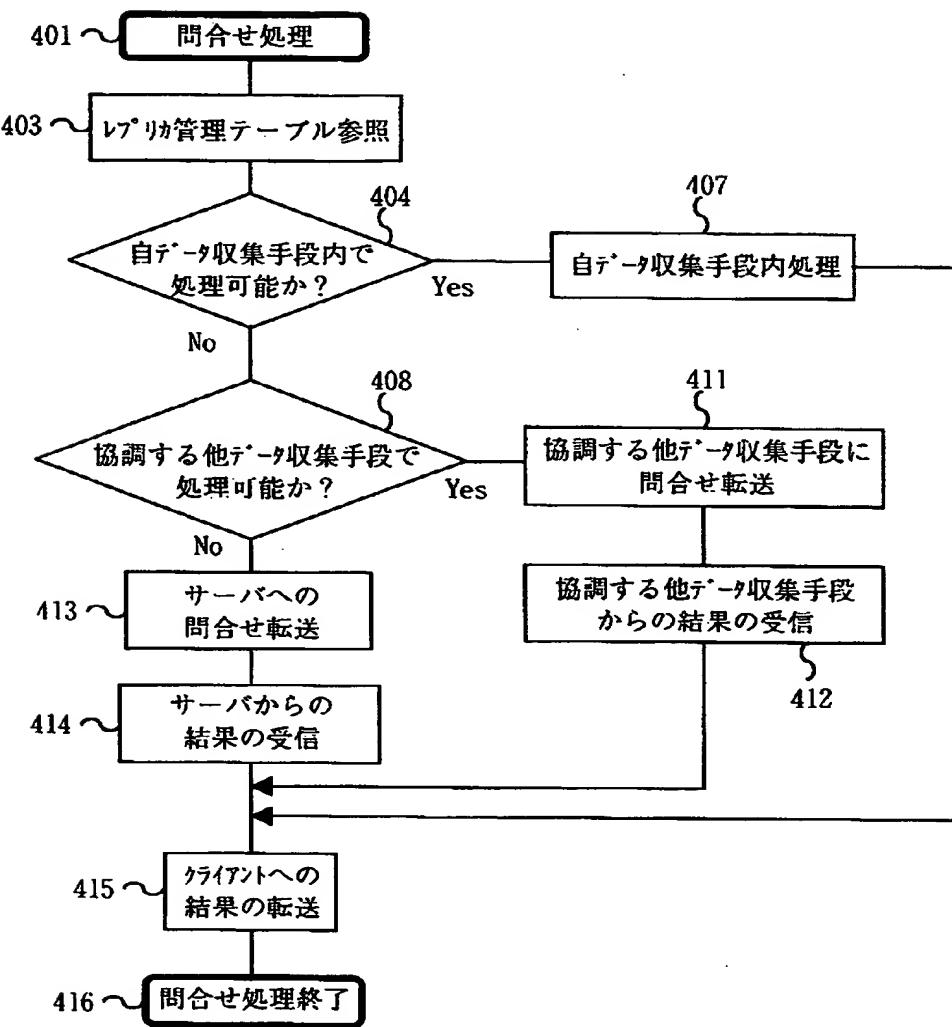
【図3】

図 3



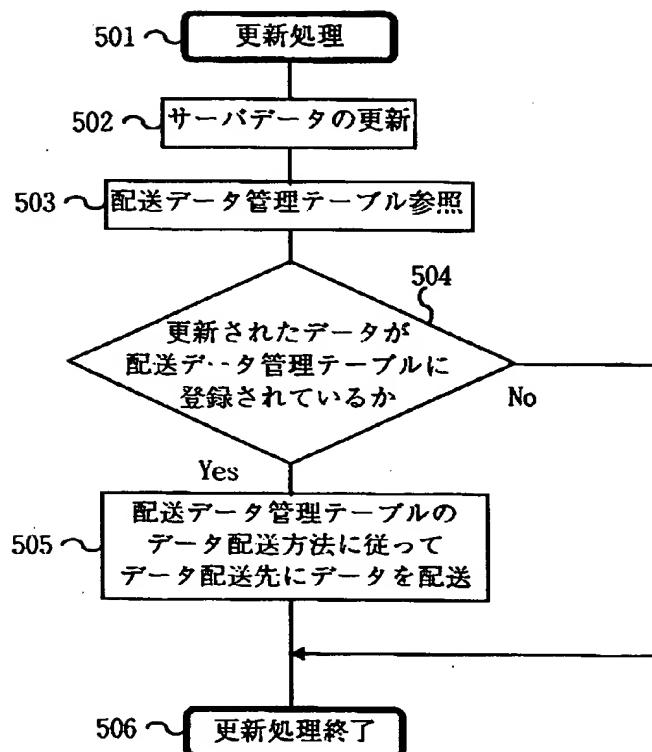
【図4】

図4



【図5】

図5



【図6】

図6

レプリカ作成要求

データ領域条件	データ品質条件	データ配送条件
注文表(注文番号, 価格, 顧客番号), 価格 ≥ 10000	-	13:00, PUSH
注文表(注文番号, 価格, 顧客番号), $8000 \geq \text{価格} \geq 5000$	fresh(注文表, 1時間) ～ 605	{1:00, 13:00}, PULL
注文表(注文番号, 価格, 顧客番号), 価格 ≤ 2000	sample(注文表, 注文番号, 10%) ～ 606	21:00～23:00の間に1度, PULL
明細表(注文番号, 発送日, 発注量), 発送日 $\geq 1990/01/01$	top-nth(明細表, 発注量, 100) ～ 607	1時間に1度, PULL
...

【図7】

図7

データタイプ	精度調整方法
関係データベースのレコード	レコードのサンプリング、フィールド切り出し
文書	要約生成、キーワード切り出し
静止画像	(非)可逆圧縮、輪郭抽出、色数削減、解像度低減、サイズ縮小
動画像	フレーム数削減、フレーム内画像圧縮
音声	音質調整、文字データへの変換

【図8】

図8

レプリカ管理テーブル

レプリカ記述 ～ 803	データ領域条件 ～ 801	データ品質条件	レプリカ位置情報	サーバ位置情報	データ配送条件
注文表(注文番号, 価格, 顧客番号), 価格 ≥ 10000	-	-	データ収集手段1	サーバ1	13:00, PUSH
注文表(注文番号, 価格, 顧客番号), 価格 ≥ 10000	807	808	データ収集手段3	サーバ1	{1:00, 13:00}, PUSH
注文表(注文番号, 価格), 価格 ≤ 3000	-	-	データ収集手段2	サーバ2	12:00, PUSH
注文表(注文番号, 価格), $3000 \leq \text{価格} \leq 5000$	sample(注文表, 注文番号, 10%)	-	データ収集手段2	サーバ2	2時間に1度, PULL

【図9】

図9

配達データ管理テーブル

901	902	903	904	
データ領域条件	データ品質条件	データ配達先	データ配達条件	
注文表(注文番号, 価格, 顧客番号), 価格 \geq 10000	-	データ収集手段 1	13:00, PUSH	905
注文表(注文番号, 価格, 顧客番号), 価格 \geq 10000	-	データ収集手段 3	11:00, 13:00, PUSH	906
注文表(注文番号, 価格, 顧客番号), 価格 \geq 50000	top-nth(注文表, 価格, 10)	データ収集手段 7	1時間に1度, PULL	
...	

挿入データ

注文番号 = 10005, 価格 = 12500, 顧客番号 = 256 ～907

【図10】

図10

レプリカ作成要求

(A)

データ領域条件	データ品質条件	データ配達条件	
注文表(注文番号, 価格, 顧客番号), 価格 \geq 10000	-	13:00, PUSH	～1001
注文表(注文番号, 価格, 顧客番号), 8000 \geq 価格 \geq 5000	fresh(注文表, 1時間)	11:00, 13:00, PUSH	
明細表(注文番号, 発送日, 発注量), 発送日 \geq 1990/01/01	-	1時間に1度, PULL	～1002
...	

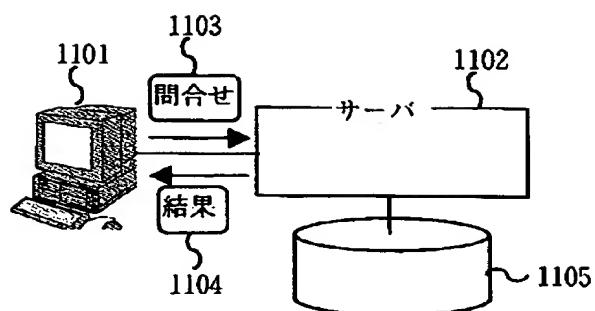
サーバ応答

(B)

要求応答	データ領域条件	データ品質条件	データ配達条件	
受け付け	-	-	13:00, PUSH	～1003
受け付け	-	-	11:00, 13:00, PUSH	
条件付き受け付け	明細表(注文番号, 発送日, 発注量), 発送日 \geq 1994/01/01	-	2時間に1度, PULL	～1004
	

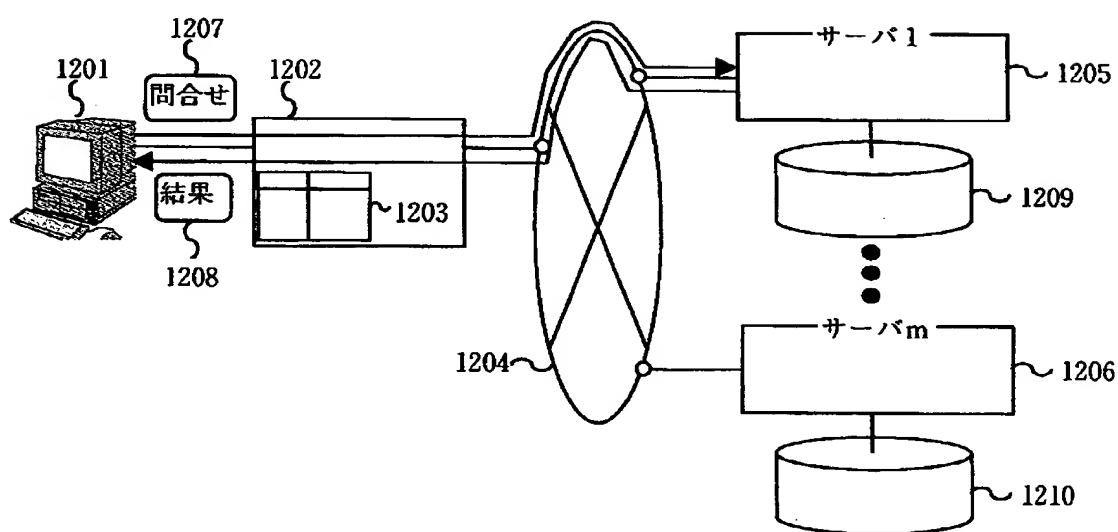
【図11】

図11



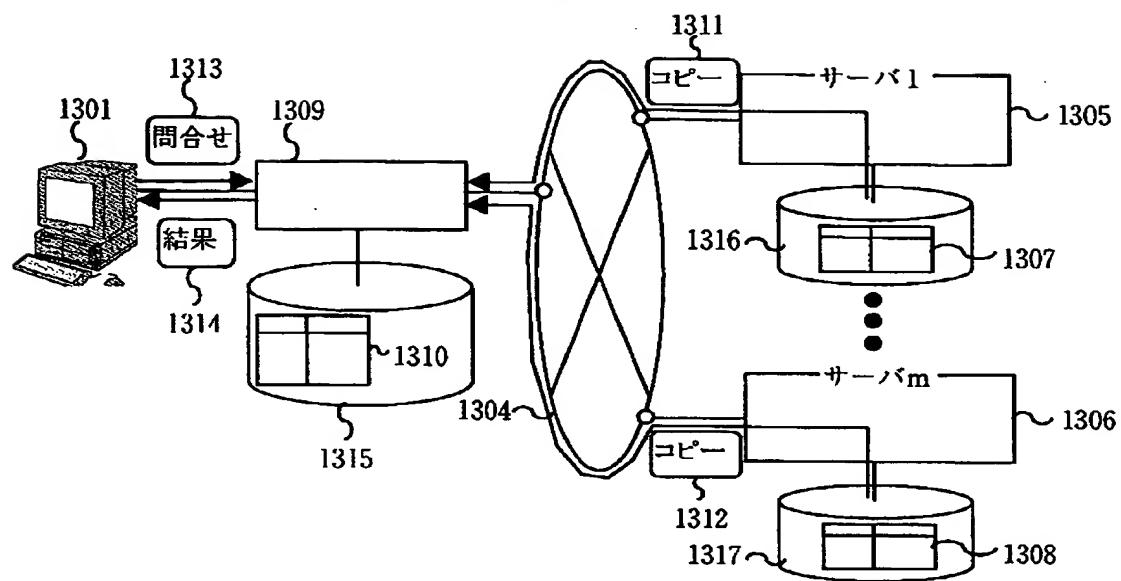
【図12】

図12



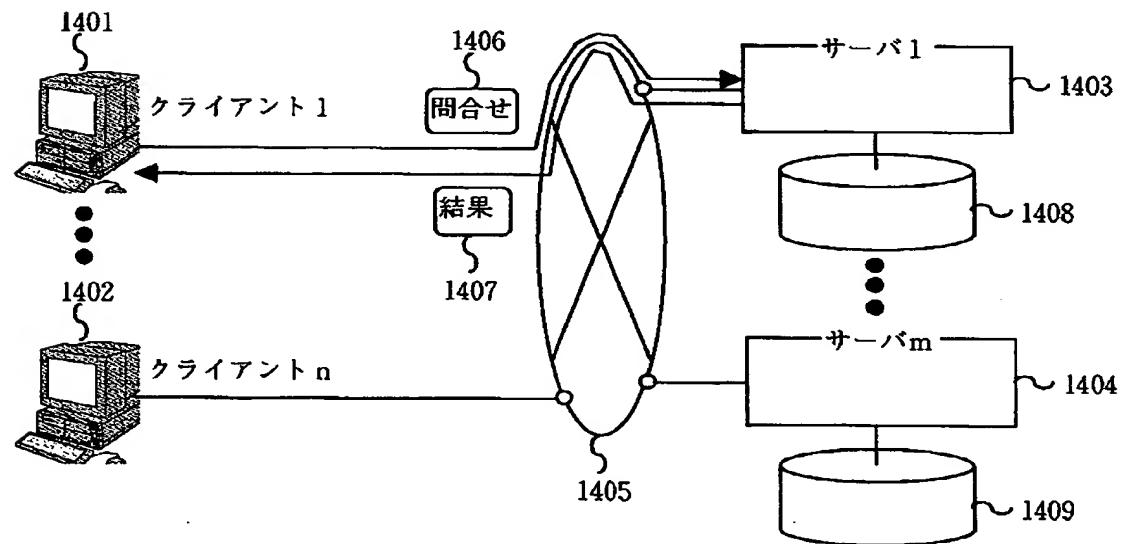
【図13】

図13



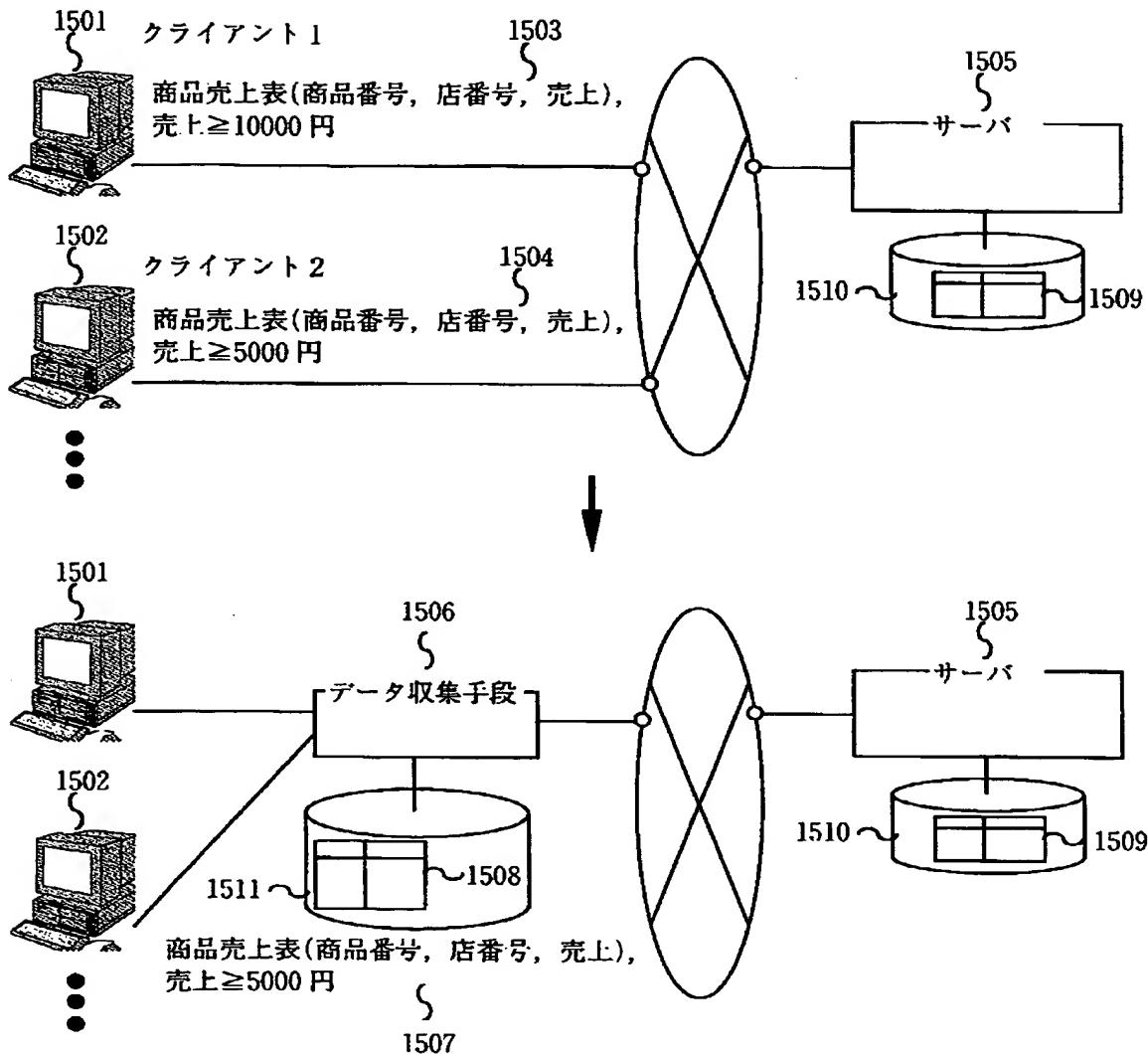
【図14】

図14



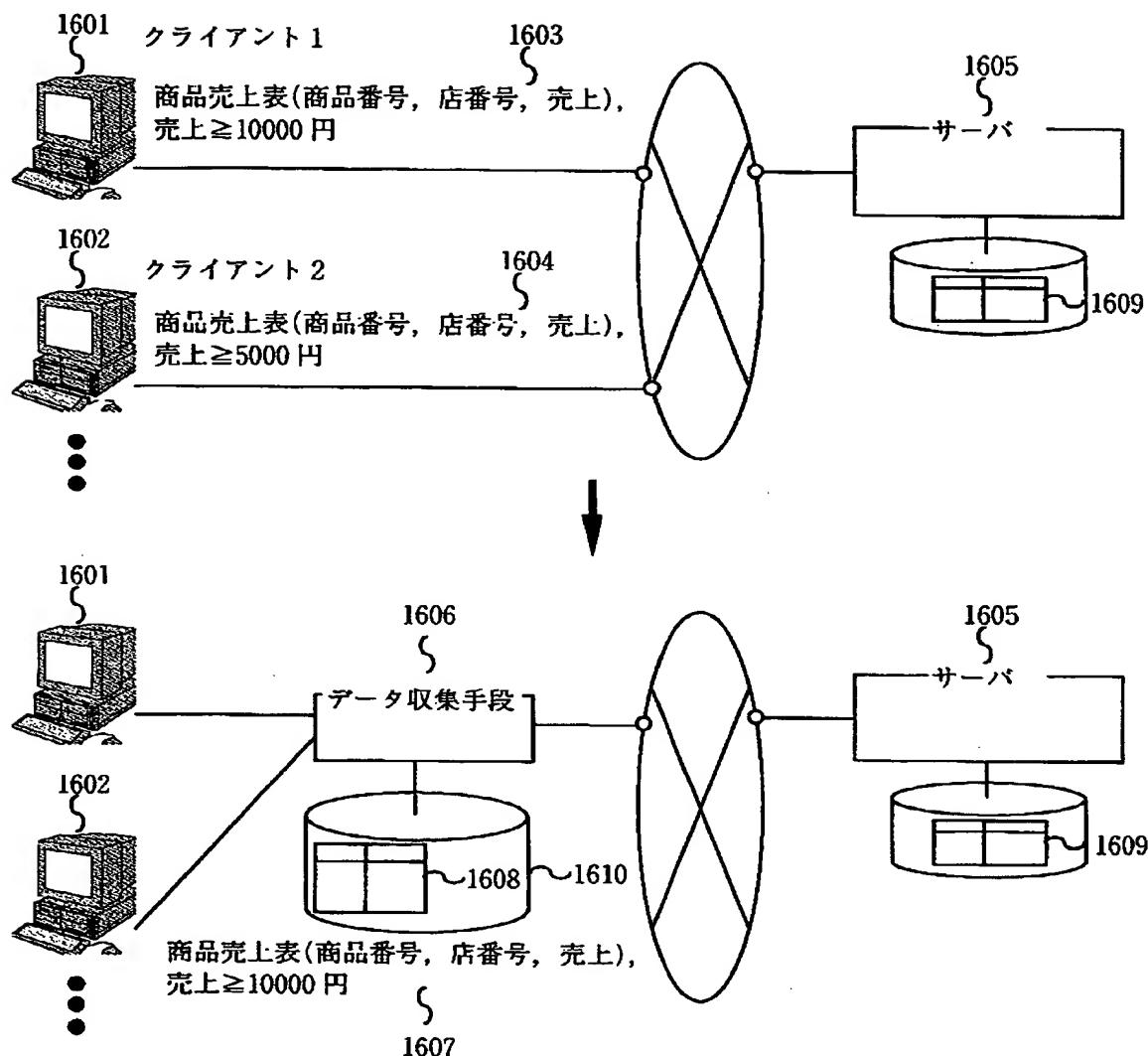
【図15】

図15



【図16】

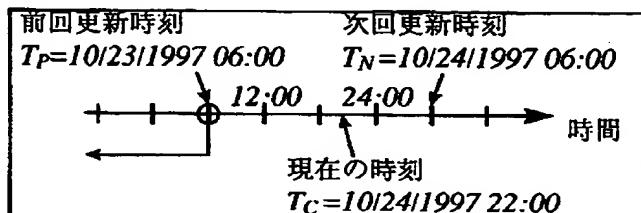
図16



【図17】

図17

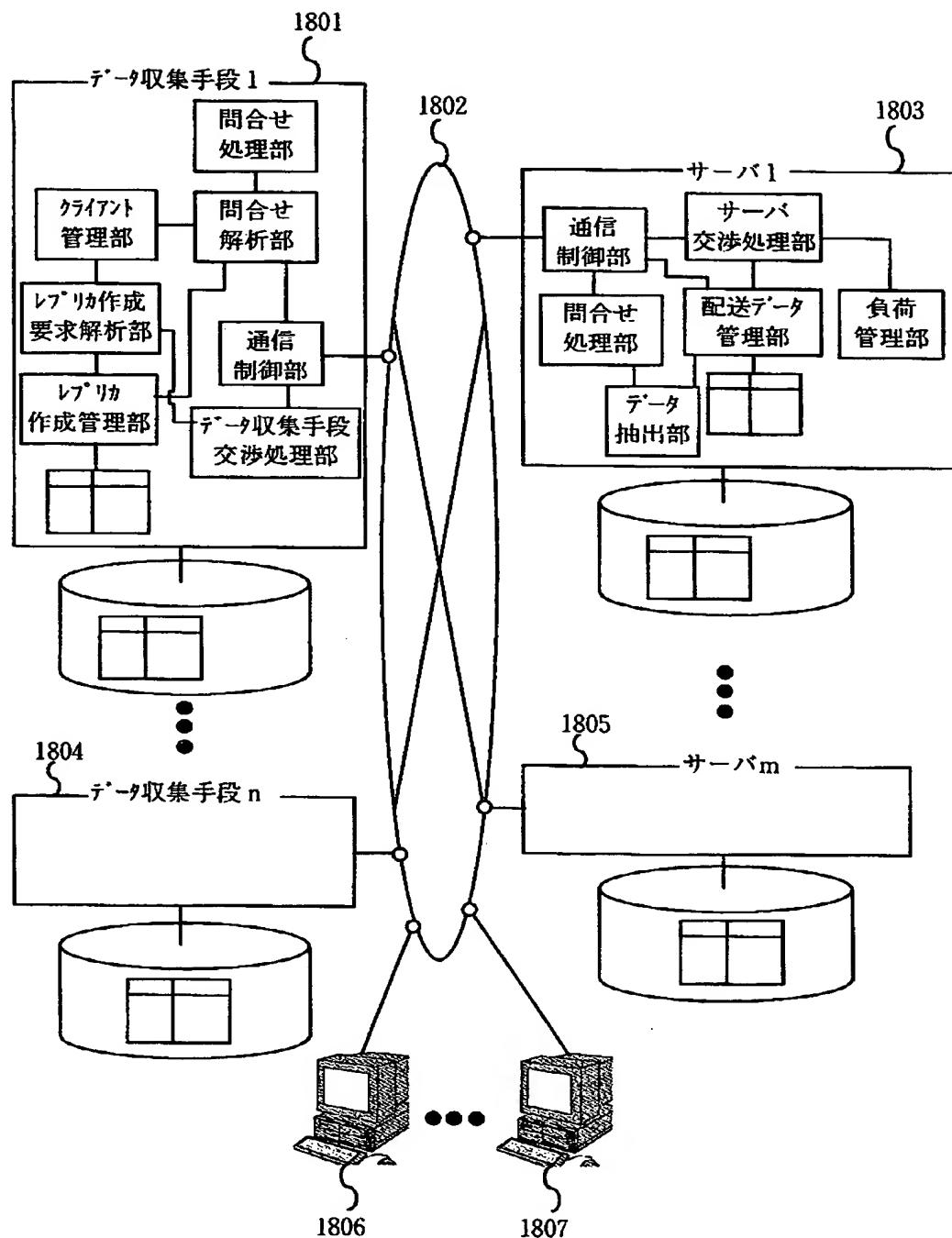
鮮度条件を用いた更新頻度の削減

鮮度条件 $T_F = 1$ 日

$$d_A(T_C) = d_A(T_P) \quad (T_C - T_P < T_F)$$

【図18】

図18



【図19】

図19

データ領域条件	データ品質条件	データ配信条件
注文表(注文番号, 価格, 顧客番号), 価格 ≥ 20000	-	11:00~15:00 の間に 1度, PUSH ～1901
注文表(注文番号, 価格, 顧客番号), 価格 ≤ 2000	-	1日に1度, PUSH ～1902
注文表(注文番号, 価格, 顧客番号), 価格 ≤ 2000	sample(注文表, 注文 番号, 10%)	21:00~23:00 の間に 1度, PULL
明細表(注文番号, 発送日, 発注量), 発送日 $\geq 1990/01/01$	top-nth(明細表, 発注 量, 100)	1時間に1度, PULL
...

【書類名】 要約書

【要約】

【課題】 分散環境において、サーバ及びネットワークの負荷削減、クライアント側の記憶装置容量の削減、および更新処理コストの削減を実現し、ネットワークを利用した実用的なデータウェアハウスシステムを提供する。

【解決手段】 クライアント（103、104）のためのサーバデータのレプリカを作成・管理するデータ収集手段（101、102）をネットワーク（113）に接続する。データ収集手段ではユーザからクライアントを介して与えられたデータの要求条件を保持し、記憶装置、C P Uなどの該データ収集手段が利用可能な資源に基づいて、例えば、データを供給するサーバ（114、122）と交渉することによって、作成するレプリカのデータ範囲および質を調整することなどが可能になる。問合せ処理の際には、クライアントからの問合せの一部あるいは全部をデータ収集手段内に作成されたレプリカ（123）を用いて処理する。レプリカを用いて処理できない問合せはサーバに転送する。

【効果】 データ収集手段で作成したレプリカを用いてクライアントからの問合せを処理することによってサーバ負荷を削減でき、実用的なデータウェアハウスシステムの構築が可能となる。

【選択図】 図1

【書類名】 職権訂正データ
【訂正書類】 特許願

〈認定情報・付加情報〉

【特許出願人】

【識別番号】 000005108

【住所又は居所】 東京都千代田区神田駿河台四丁目6番地

【氏名又は名称】 株式会社日立製作所

【代理人】 申請人

【識別番号】 100068504

【住所又は居所】 東京都千代田区丸の内1-5-1 株式会社日立製作所 知的所有権本部内

【氏名又は名称】 小川 勝男

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所